

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-175091

(P2002-175091A)

(43) 公開日 平成14年6月21日 (2002. 6. 21)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 1 0 L 13/00		A 6 3 H 11/00	Z 2 C 1 5 0
A 6 3 H 11/00		B 2 5 J 5/00	C 3 C 0 0 7
B 2 5 J 5/00		13/00	Z 5 D 0 4 5
13/00		G 0 6 F 3/16	3 3 0 K
G 0 6 F 3/16	3 3 0		3 4 0 N

審査請求 未請求 請求項の数18 O L (全 20 頁) 最終頁に続く

(21) 出願番号 特願2000-372091 (P2000-372091)

(22) 出願日 平成12年12月6日 (2000. 12. 6)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 佐部 浩太郎

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(72) 発明者 ビエール イブス オードイエ

フランス国、75005 パリ市 アミヨ通り
6番地 ソニーコンピュータサイエンス研究所内

(74) 代理人 100067736

弁理士 小池 晃 (外2名)

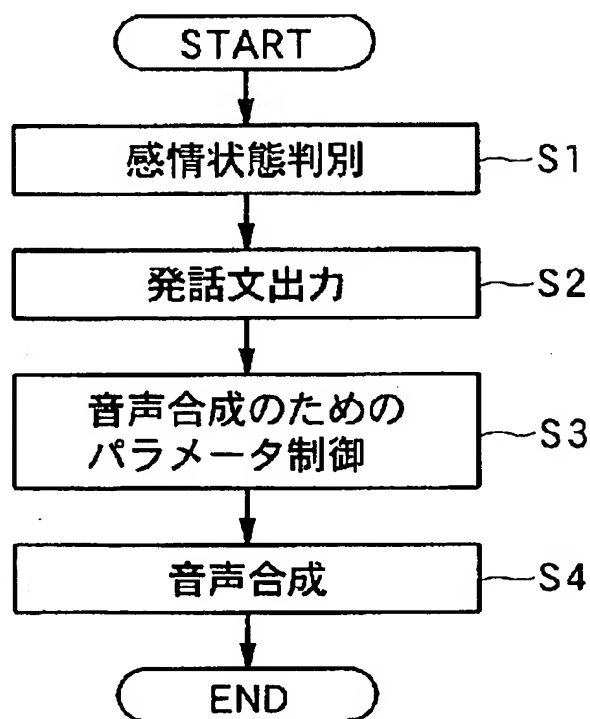
最終頁に続く

(54) 【発明の名称】 音声合成方法及び装置、並びにロボット装置

(57) 【要約】

【課題】 生物等の感情表現により近いような聴覚的な感情表現をする。

【解決手段】 ロボット装置は、感情モデルの感情状態を判別する感情判別工程（ステップS1）と、音声として発する内容を表す発話文を出力する発話文出力工程（ステップS2）と、感情判別工程により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御工程（ステップS3）と、発話文出力工程により出力された発話文を音声合成部に入力して制御されたパラメータに基づいて音声合成する音声合成工程（ステップS4）とにより、音声合成による発話文の生成を行う。



【特許請求の範囲】

【請求項 1】 少なくとも感情モデルを有する発音主体からの情報に基づいて音声を作成する音声合成方法であって、
上記発音主体の上記感情モデルの感情状態を判別する感情判別工程と、
音声として発する内容を表す発話文を出力する発話文出力工程と、
上記感情判別工程により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御工程と、
上記発話文出力工程により出力された発話文を音声合成部に入力して上記制御されたパラメータに基づいて音声合成する音声合成工程とを有することを特徴とする音声合成方法。

【請求項 2】 上記発話文は、無意味な内容の文であることを特徴とする請求項 1 記載の音声合成方法。

【請求項 3】 上記発話文出力工程は、上記感情モデルの感情状態が所定の閾値を超えたときに上記発話文を出力して上記音声合成部に供給することを特徴とする請求項 1 記載の音声合成方法。

【請求項 4】 上記発話文出力工程は、発話毎にランダムに得られた上記発話文を出力して上記音声合成部に供給することを特徴とする請求項 1 記載の音声合成方法。

【請求項 5】 上記発話文は、複数の音素を有して成り、上記パラメータは、上記音素の持続時間、ピッチ、音量を含むことを特徴とする請求項 1 記載の音声合成方法。

【請求項 6】 上記発音主体は、供給された入力情報に基づいて動作を行う自律型ロボット装置であり、上記感情モデルとして、上記動作に起因する感情モデルを有し、上記入力情報に基づいて上記感情モデルの状態を変化させることにより上記動作を決定する感情モデル変化工程を有することを特徴とする請求項 1 記載の音声合成方法。

【請求項 7】 少なくとも感情モデルを有する発音主体からの情報に基づいて音声を作成する音声合成装置であって、
上記発音主体の上記感情モデルの感情状態を判別する感情判別手段と、
音声として発する内容を表す発話文を出力する発話文出力手段と、
上記感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、
上記発話文出力手段により出力された発話文が供給され、上記制御されたパラメータに基づいて音声合成する音声合成手段とを有することを特徴とする音声合成装置。

【請求項 8】 上記発話文は、無意味な内容の文である

ことを特徴とする請求項 7 記載の音声合成装置。

【請求項 9】 上記発話文出力手段は、上記感情モデルの感情状態が所定の閾値を超えたときに上記発話文を出力して上記音声合成手段に供給することを特徴とする請求項 7 記載の音声合成装置。

【請求項 10】 上記発話文出力手段は、発話毎にランダムに得られた上記発話文を出力して上記音声合成手段に供給することを特徴とする請求項 7 記載の音声合成装置。

【請求項 11】 上記発話文は、複数の音素を有して成り、上記パラメータは、上記音素の持続時間、ピッチ、音量を含むことを特徴とする請求項 7 記載の音声合成装置。

【請求項 12】 上記発音主体は、供給された入力情報に応じた動作を行う自律型ロボット装置であり、上記感情モデルとして、上記動作に起因する感情モデルを有し、上記入力情報に基づいて上記感情モデルの状態を変化させることにより上記動作を決定する感情モデル変化手段を備えることを特徴とする請求項 7 記載の音声合成装置。

【請求項 13】 供給された入力情報に基づいて動作を行う自律型のロボット装置であって、
上記動作に起因する感情モデルと、
上記感情モデルの感情状態を判別する感情判別手段と、
音声として発する内容を表す発話文を出力する発話文出力手段と、
上記感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、
上記発話文出力手段により出力された発話文が供給され、上記制御されたパラメータに基づいて音声合成する音声合成手段とを備えることを特徴とするロボット装置。

【請求項 14】 上記発話文は、無意味な内容の文であることを特徴とする請求項 13 記載のロボット装置。

【請求項 15】 上記発話文出力手段は、上記感情モデルの感情状態が所定の閾値を超えたときに上記発話文を出力して上記音声合成手段に供給することを特徴とする請求項 13 記載のロボット装置。

【請求項 16】 上記発話文出力手段は、発話毎にランダムに得られた上記発話文を出力して上記音声合成手段に供給することを特徴とする請求項 13 記載のロボット装置。

【請求項 17】 上記発話文は、複数の音素を有して成り、上記パラメータは、上記音素の持続時間、ピッチ、音量を含むことを特徴とする請求項 13 記載のロボット装置。

【請求項 18】 上記入力情報に基づいて上記感情モデルの状態を変化させることにより上記動作を決定する感情モデル変化手段を備えることを特徴とする請求項 13

記載のロボット装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、発話主体が出力する音声を生産するための音声合成方法及び音声合成装置、並びに音声を出力するロボット装置に関する。

【0002】

【従来の技術】近年、外観形状が犬や猫等の動物に模して形成された例えばペット型のロボット装置が提供されている。そのようなロボット装置には、外部からの情報や内部の状態に応じて自律的に動作するものがある。

【0003】そのような、ロボット装置に用いられる人工知能(AI:artificial intelligence)は、推論・判断等の知的な機能を人工的に実現したものであり、さらに感情や本能等の機能をも人工的に実現することが試みられている。このような人工知能の外部への表現手段としての視覚的な表現手段や聴覚的な表現手段等の内で、聴覚的なものの一例として、音声を用いることが挙げられる。

【0004】そのようなロボット装置においては、その発声を使って人間(飼主等)に自身の感情を訴える機能は有効である。これは、人間は実際の犬や猫等のペットが何を言っているのか直接理解することはできないが、経験から愛犬、愛猫が今どのような機嫌なのか知ることができるのであり、そのひとつの要素がペットの発声により判断しているからである。

【0005】

【発明が解決しようとする課題】ところで、現在市場に供給されているロボット装置として、聴覚的な感情表現を電子的な音(電子音)で行っているものが知られている。具体的には、例えば、短く高い音が喜びを表し、ゆっくりとした低い音が悲しみを表すなどである。そして、これらの電子音は予め作曲されており、人間の主観によってそれぞれの感情クラスに振り分けられ、再生に用いられる。ここで、感情クラスとは、喜び、怒り等として分類される情動のクラスである。従来よりなされている電子音を用いた聴覚的な感情表現においては、

(i) 機械的である、(ii) 常に同じ表現を繰り返す、(iii) 表現力が適切であるか不明である、などの点が犬や猫等の生物のペットの感情表現と大きく異なる点として挙げられ、さらなる改善が望まれているのが実情である。

【0006】そこで、本発明は、上述の実情に鑑みてなされたものであり、生物等の感情表現により近いような聴覚的な感情表現を可能とする音声合成方法及び音声合成装置、並びにロボット装置の提供を目的とする。

【0007】

【課題を解決するための手段】本発明に係る音声合成方法は、上述の課題を解決するために、発音主体の感情モデルの感情状態を判別する感情判別工程と、音声として

発する内容を表す発話文を出力する発話文出力工程と、感情判別工程により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御工程と、発話文出力工程により出力された発話文を音声合成部に入力して制御されたパラメータに基づいて音声合成する音声合成工程とを有する。

【0008】このような音声合成方法は、発声主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生産する。

【0009】また、本発明に係る音声合成装置は、上述の課題を解決するために、発音主体の感情モデルの感情状態を判別する感情判別手段と、音声として発する内容を表す発話文を出力する発話文出力手段と、感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成する音声合成手段とを備える。

【0010】このような構成を備える音声合成装置は、発音主体の感情モデルの感情状態を判別する感情判別手段により判別された感情状態に応じて音声合成のためのパラメータをパラメータ制御手段により制御して、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成手段により音声合成する。これにより、音声合成装置は、発声主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生産する。

【0011】また、本発明に係るロボット装置は、上述の課題を解決するために、動作に起因する感情モデルと、感情モデルの感情状態を判別する感情判別手段と、音声として発する内容を表す発話文を出力する発話文出力手段と、感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成する音声合成手段とを備える。

【0012】このような構成を備えるロボット装置は、動作に起因する感情モデルの感情状態を判別する感情判別手段により判別された感情状態に応じて音声合成のためのパラメータをパラメータ制御手段により制御して、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成手段により音声合成する。これにより、ロボット装置は、発声主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生産する。

【0013】

【発明の実施の形態】先ず、本発明に係る音声合成方法及び装置、並びにロボット装置の好ましい実施の形態の説明に先立って、音声による感情表現をペット型のロボット装置等の機能として採用することの意味、及び適切

な音声による感情表現について説明する。

【0014】(1) 音声による感情表現

ロボット装置等における機能として発話に感情表現を加えることは、人間との親密性を高めるために非常に有効に働く。また、単に社交性の向上だけではなく、自身の満足や不満足を示すことで人間に対して刺激を要求することができる。このような機能は、学習機能を持つロボット装置においては有効に作用する機能になる。

【0015】一方、人間の持つ感情と発せられる音声の音響特性に相関があるかどうかについて、Fairbanksの報告(Fairbanks G.(1940), Recent experimental investigations of vocal pitch in speech, Journal of the Acoustical Society of America, (11):457-466.)やBurkhardtらの報告(Burkhardt F. and Sendlmeier W. F., Verification of Acoustic Correlates of Emotional Speech using Formant Synthesis, ISCA Workshop on Speech and Emotion, Belfast 2000.)等があり、このように多くの研究者による報告が知られている。

【0016】これらの報告によると、発話は心理学的な条件や幾つかの基本的な情動クラスとの相関があることが判っている。また逆に驚きと怖れ、退屈と悲しみなどのある特定の感情に関しては違いを見つけることが難しいことなども報告されている。ある情動についてはある身体的な状態と結びついており、容易に予想可能な物理的な影響を発話に対してもたらす。

【0017】例えば、ある人が怒りや怖れや喜びを覚えたときは交感神経系が奮起し、心拍数や血圧が上昇する。口内は乾き、時には筋肉に震えが起こる。こうしたときには発話は大きく速くなり、高周波成分に強いエネルギーを持つことになる。またある人が退屈や悲しみを感じているとき、副交感神経系が奮起する。心拍数や血圧は減少し、唾液が多く分泌される。この結果発話はゆっくりとしたピッチの遅いものになる。そして、これら身体的特徴は万国共通であるために、基本的な情動と発話の音響的な特性の間には民族や文化に寄らない相関関係が存在すると考えられている。

【0018】また、日本人とアメリカ人に意味の無い単語の羅列を色々な感情で発話させてお互いの感情がどの程度認識されるかの実験を行った結果がAbelinらの報告(Abelin A., Allwood J., Cross Linguistic Interpretation of Emotional Prosody, Workshop on Emotions in Speech, ISCA Workshop on Speech and Emotion, Belfast 2000.)やTickleの報告(Tickle A., English and Japanese Speaker's Emotion Vocalizations and Recognition: A Comparison Highlighting Vowel Quality, SCA Workshop on Speech and Emotion, Belfast 2000.)にある。このような報告にある結果から、(i)言語の違いによって感情の認識率は変わらない、(ii)認識結果はあまり芳しくなく60%程度である、という、2つの点が明らかにされている。

【0019】これらの研究結果を考慮すると、人間と例えばロボット装置との間に、意味の伝達を必要としない無意味語による感情の伝達が、非常にあいまいではあるが、可能であることがわかる。例えば、その感情の認識率は60%程度とされている。また、情動と音響特性の相関をモデル化することでそうした発話を合成することが可能であることも示している。

【0020】本発明の実施の形態では、このような音響特性に基づいた発話をするることにより、感情を表出させて発話をするようにしている。さらに、本発明の実施の形態では、(i)スピーチのような、(ii)意味を持たない、(iii)毎回違う、発声を実現している。

【0021】ここで、図1は、本発明に係る音声合成方法の実施の形態の基本構成を示すフローチャートであり、発音主体としては、少なくとも感情モデル、音声合成手段及び発音手段を有する例えばロボット装置を想定しているが、これに限定されず、各種ロボット装置や、ロボット以外の各種コンピュータAI (artificial intelligence) 等への適用も可能であることは勿論である。なお、感情モデルについては後で詳述する。

【0022】この図1において、最初のステップS1では、発音主体の感情モデルの感情状態を判別する。具体的には、例えば、周囲の環境(外的要因)や内部の状態(内的要因)に応じて感情モデルの状態(感情状態)が変化し、この感情状態について、例えば平静、怒り、悲しみ、喜び、落ち着き、のいずれかであるかを判別する。

【0023】ここで、ロボット装置の場合には、行動モデルとして、内部に確率状態遷移モデル(例えば、後述するように、状態遷移表を有するモデル)を有しており、各状態が認識結果や感情や本能の値によって異なる遷移確率テーブルを持ち、その確率に従って次の状態へ遷移し、この遷移に関連づけられた行動を出力する。

【0024】感情による喜びや悲しみの表現行動がこの確率状態遷移モデル(或いは確率遷移表)に記述されており、この表現行動の1つとして、音声による(発話による)感情表現が含まれている。従って、この具体例では、感情モデルの感情状態を表すパラメータを行動モデルが参照することにより決定される行動の1つの要素として感情表現があり、行動決定部の一部機能として、感情状態の判別が行われることになる。

【0025】なお、本発明は、この具体例に限定されるものではなく、ステップS1では少なくとも感情モデルの感情状態の判別が行われればよく、後のステップでは、判別された感情状態を音声で表現するような音声合成が行われるものである。

【0026】次のステップS2では、音声として発する内容を表す発話文を出力する。このステップS2は、ステップS1の前、あるいは後述するステップS3の後でもよい。また、毎回新たな発話文を生成してもよく、予

め生成されて用意された複数の発話文のいずれかをランダムに選択するようにしてもよい。ただし、本発明の実施の形態においては、無意味な発話文であることが必要とされる。これは、特にロボット装置に適用する場合において、実際に意味を持った対話を実現することは困難であるが、無意味語の発話であれば簡単な構成で実現でき、しかも感情表現が伴うことにより、無意味語でも充分に対話しているように感じさせることができ、また、無意味語の方が聞き手の側の想像の余地が広がって、不適切な意味のある発話内容よりも親近感、親密感が高まること等を考慮したものである。また、発話文の生成あるいは選択をランダムに行うことにより、音声合成されて再生される発声が毎回違うようになり、常に新鮮に感じるようになる。

【0027】このようにこのステップS2において出力される発話文は、ランダムな単語で構成される文であり、具体的には、それぞれの単語をランダムな音節にすることによって実現している。ここでいう音節とは、音素である子音Cと母音Vとを組み合わせ、例えばCV若しくはCCVとしたものである。実施の形態では、音素を予め用意しており、その全ての音素は、最初は固定されたパラメータとしての持続時間やピッチを有しているが、感情状態の判別結果に応じて制御される。感情状態の判別結果に応じたそのようなパラメータの制御により、感情表現がなされた発話になる。このような感情表現の判別結果に応じたパラメータの制御については後で詳述する。

【0028】なお、本実施の形態においては、出力される発話文は、感情モデルの感情状態やその判別結果とは無関係であるが、感情状態等に応じて、出力される発話文をある程度調整したり、発話文の生成あるいは選択処理自体を制御するようにしてもよい。

【0029】次に、ステップS3では、上述のステップS1での感情状態の判別結果に応じて、音声合成のためのパラメータを制御する。音声合成のためのパラメータとは、例えば上述の音素の持続時間やピッチ、あるいは音量等のことであり、感情状態の判別結果、例えば、平静、怒り、悲しみ、喜び、落ち着き等に応じてこれらのパラメータを変えることで感情表現を行わせる。具体的には、上述の判別結果としての各感情（平静、怒り、悲しみ、喜び、落ち着き等）に対応するパラメータの組合せテーブルを予め作成しておき、実際に判別された感情に応じてこれらのテーブルを切り換えることが挙げられる。各感情に応じて用意するテーブルについては後で具体例を示している。

【0030】次のステップS4では、上述のステップS2で出力された発話文を音声合成器（スピーチシンセサイザ：speech synthesizer）に送り、上述のステップS3で制御されたパラメータに従って音声合成する。音声合成されて得られた音声時系列データは、D/A変換器

やアンプ等を介してスピーカに送られることにより、実際の音声として発せられる。例えば、ロボット装置である場合には、このような処理が、いわゆるバーチャルロボットにおいてなされてスピーカから、そのときの感情を表現するような発話がなされるようになる。

【0031】以上説明した本発明の基本的な実施の形態によれば、身体的状態に関連する感情に応じて、音声合成のためのパラメータ（音素の持続期間、ピッチ、音量等）を制御することで、感情表現をともなった発話ができ、音素をランダムに選ぶために単語や文自体に意味を持たせる必要がなく、それでいて発話のように聞こえ、また、前述パラメータの一部をランダムに変更したり、音素の組合せや、単語、文の長さをランダムに決定することで合成する度に違う発話を作り出すことができると共に、制御するパラメータが少ないために実装が簡便である。

【0032】（2）感情と無意味語の合成アルゴリズム感情と発話文となる無意味語との合成アルゴリズムについて詳細に説明する。実施の形態において、合成アルゴリズムの目標とするところは、スピーチのような、意味を持たない、毎回違う、発話文を生成することである。そして、そのような発話文に感情表現を伴わせることである。

【0033】このような発話文の生成のために音声合成器（スピーチシンセサイザ：speech synthesizer）、あるいは音声合成システムを利用する。音声合成器（スピーチシンセサイザ：speech synthesizer）、あるいは音声合成システムへの入力は、音素のリストとそれぞれの音素に対する、持続期間、目標とするピッチと到達時刻（持続期間に対するパーセンテージで表現）等である。このような音声合成を実現するアルゴリズムの概略は、以下のようになっている。

【0034】（2-2）発話文の生成無意味語の発話文の生成は、ランダムな単語により構成することに実現している。さらに、それぞれの単語をランダムな音節によって構成している。ここで、音節は、音素である子音Cと母音Vとを組み合わせたり、CVもしくはCCVとして表されるものとしている。ここで、音素は、リストとして有している。そして、リストとして有するすべての音素は、固定された持続期間やピッチをもって先ず登録されている。

【0035】例えば、ある音素「b」は「448 10 150 8 0 158」といった値によって表現されて、リストに登録されている。ここで、「448」は、音素「b」の持続期間が448msであることを示している。また、次の「10」及びその次の「150」は、持続時間448msの10%で150Hzに到達することを示している。また、その次の「80」及びその次の「158」は、持続時間448msの80%には158Hzに到達することを示している。このようにしてリストのすべての音素が表現される。

【0036】図2には、「131 80 179」により与えられる音素「b」と、「77 20 200 80 229」により与えられる音素「@」と、「405 80 169」により与えられる音素「b」との結合によって表現される音節を示している。この例では、実際には不連続の関係にある各音素を連続となるように結合した音節として示している。

【0037】このような音節を構成するものとなる音素が各感情表現にあわせて変更を加えられることで、発話文に感情表現がなされるようになる。具体的には、上述したような音素の個性或いは特性を示す情報でもある持続期間やピッチが感情表現のために変更される。

【0038】このような音素から構成する発話文は、大別して、単語の組合せであり、また、その単語が、音節の組合せとされ、さらに、その音節が音素の組合せとされて構成されるものである。以下にそのような発話文を構成する各段階毎<1>～<5>における処理について詳しく説明する。

<1> 先ず、文の中の単語の数を決める。例えば、20～MAXWORDSの間の乱数として決定する。MAXWORDSは、文章を構成する単語の最大数であり、音声合成のためのパラメータである。

<2> 各単語を生成する。具体的には、先ず、文章において単語にアクセントがあるかどうかを確率（PROB ACCENT）で決定する。

【0039】以下の<3-1>～<3-7>の手続きにより単語の音節及びその音素を決定し、単語を決定する。

<3-1> 各単語の中の音節の数を決める。例えば、2～MAXSYLLの間の乱数として決定する。MAXSYLLは、単語を構成する音節の最大数であり、音声合成のためのパラメータである。

<3-2> ここで、アクセントがある単語であるならば、音節のうちの一つをランダムに選んでアクセントをマークする。

<3-3> 各音節をC V表現かC C V表現のものとして決定する。例えば、C V表現の音節が0.8%の確率で選択されるように決定する。

<3-4> そのように選んだC V或いはC C VのC及びVに割り当てる子音と母音を音素データベース（或いは音素リスト）からランダムに読み出す。

<3-5> 各音素の持続期間を、MEANDUR+random(DURVAR)で計算する。ここで、MEANDURは固定の持続時間であり、random(DURVAR)は乱数に決定される値である。ここで、MEANDUR及びDURVARは、音声合成のためのパラメータである。

<3-6-1> 音素のピッチの計算を、 $e = \text{MEANPITCH} + \text{random}(\text{PITCHVAR})$ で計算する。ここで、MEANPITCHは固定のピッチであり、random(PITCHVAR)は乱数に決定される値である。ここで、MEANPITCH及びPITCHVARは、例えば感情に応じて決定されるパラメータにしている。

<3-6-2> ここで、音素が子音なら、 $e - \text{PITCHVAR}$ として子音のピッチを得る。また、音素が母音なら、 $e + \text{PITCHVAR}$ として母音のピッチを得る。

<3-7-1> もし、音節にアクセントがあるなら、持続期間にDURVARを追加する。

<3-7-2> そして、アクセントがある場合において、DEFAULTCONTOUR=risingであるときには、子音のピッチを、 $\text{MAXPITCH} - \text{PITCHVAR}$ とし、母音のピッチを、 $\text{MAXPITCH} + \text{PITCHVAR}$ とする。また、DEFAULTCONTOUR=fallingであるときには、子音のピッチを、 $\text{MAXPITCH} + \text{PITCHVAR}$ とし、母音のピッチを、 $\text{MAXPITCH} - \text{PITCHVAR}$ とする。さらに、DEFAULTCONTOUR=stableであるときには、子音及び母音のピッチを、 MAXPITCH とする。DEFAULTCONTOURは、音節の特性（CONTOUR）を示すものとなる。また、MAXPITCHは、音声合成のためのパラメータである。

【0040】以上のような<3-1>～<3-7>の手続きにより単語の音節及びその音素が決定される。そして、最後に発話文の最後にくる単語のcontour（輪郭）を変更する処理を行う。

<4-1> 文章の最後の単語にアクセントが無ければ、 $e = \text{PITCHVAR}/2$ にする。ここで、CONTOURLASTWORD=falingであるときには、各音節について、 $-(l+1)*e$ を加え、 $e = e + e$ とする。lは音素のインデックスを示すものとなる。また、パラメータCONTOURLASTWORD=risingであるときには、各音節について、 $+(l+1)*e$ を加え、 $e = e + e$ とする。

<4-2> 一方、最後の単語にアクセントがあれば、CONTOURLASTWORD=fallingであるときには、各音節の持続期間に、DURVARを加える。そして、子音のピッチを、 $\text{MAXPITCH} + \text{PITCHVAR}$ とし、母音のピッチを、 $\text{MAXPITCH} - \text{PITCHVAR}$ とする。もし、パラメータCONTOURLASTWORD=risingであるときには、各音節の持続期間に、DURVARを加える。そして、子音のピッチを、 $\text{MAXPITCH} - \text{PITCHVAR}$ とし、母音のピッチを、 $\text{MAXPITCH} + \text{PITCHVAR}$ とする。

<5> 最後に文全体のボリュームをVOLUMEに設定する。ここで、VOLUMEは、音声合成のためのパラメータである。

【0041】以上のような各段階毎における処理により、発話文が生成されるようになる。そして、発話文の決定に使用する幾つかのパラメータを乱数によって与えているので、意味を持たない、毎回違う、無意味語を生成することができるようになる。そして、上述したような各種パラメータを感情に応じて与えることにより、そのような発話文に感情表現がなされるようになる。

【0042】なお、図3及び図4には、上述したような処理をハードウェアによって実現するためのプログラム（ソースコード）の記述を示している。図3にはプログラムの前段部分、図4には後段部分を示している。

【0043】（2-2）各感情に応じて与えられるパラ

メータ

上述したような文章生成のアルゴリズムにおいて使用したパラメータを感情に応じて制御することで発話文に感情表現をもたせることができる。ここで、例えば、そのように発話文の表出される感情としては、平静、怒り、悲しみ、喜び或いは落ち着き (calm, anger, sadness, happiness, comfort) 等が挙げられる。なお、ここで列記する感情に限定されないことはいふまでもない。

【0044】例えば、このような情動はそれぞれ、奮起 (Arousal) と力価 (valence) とを要素とする特徴空間上に表現できる。例えば、図5に示すように、奮起 (Arousal) と力価 (valence) とを要素とする特徴空間上において、怒り、悲しみ、喜び或いは落ち着き (anger, sadness, happiness, comfort) の領域が構成され、その中心に平静 (calm) の領域が構成されるというようである。例えば、「怒り (anger)」は奮起とネガティブとして表され、「悲しみ (sadness)」は「奮起でないとネガティブ」として表される。

【0045】以下の表には、怒り、悲しみ、喜び及び落ち着き等の各感情に対応して予め決定されているパラメータ (少なくとも音素の持続期間 (DUR)、ピッチ (PITCH) 及び音量 (VOLUME) 等) の組合せテーブルを示している。このようなテーブルを各感情の特質に基づいて予め生成しておく。

【0046】

【表1】

平静 (Calm)	
パラメータ	状態又は値
LASTWORDACCENTED	no
MEANPITCH	280
PITCHVAR	10
MAXPITCH	370
MEANDUR	200
DURVAR	100
PROBACCENT	0.4
DEFAULTCONTOUR	rising
CONTOURLASTWORD	rising
VOLUME	1

【0047】

【表2】

怒り (Anger)	
パラメータ	状態又は値
LASTWORDACCENTED	no
MEANPITCH	450
PITCHVAR	100
MAXPITCH	500
MEANDUR	150
DURVAR	20
PROBACCENT	0.4
DEFAULTCONTOUR	falling
CONTOURLASTWORD	falling
VOLUME	2

【0048】

【表3】

悲しみ (sadness)	
パラメータ	状態又は値
LASTWORDACCENTED	nil
MEANPITCH	270
PITCHVAR	30
MAXPITCH	250
MEANDUR	300
DURVAR	100
PROBACCENT	0
DEFAULTCONTOUR	falling
CONTOURLASTWORD	falling
VOLUME	1

【0049】

【表4】

落ち着き (comfort)	
パラメータ	状態又は値
LASTWORDACCENTED	t
MEANPITCH	300
PITCHVAR	50
MAXPITCH	350
MEANDUR	300
DURVAR	150
PROBACCENT	0.2
DEFAULTCONTOUR	rising
CONTOURLASTWORD	rising
VOLUME	1

【0050】

【表5】

喜び (happiness)	
パラメータ	状態又は値
LASTWORDACCENTED	1
MEANPITCH	400
PITCHVAR	100
MAXPITCH	600
MEANDUR	170
DURVAR	50
PROBACCENT	0.3
DEFAULTCONTOUR	rising
CONTOURLASTWORD	rising
VOLUME	2

【0051】このようにして予め用意しておいた各感情に対応されるパラメータからなるテーブルを、実際に判別された感情に応じて切り換えることにより、音声合成に使用するパラメータの感情に応じた制御を実現している。

【0052】そして、そのように感情に応じて選択されたテーブルのパラメータを使用した音声合成がなされることにより、感情表現がなされた発話文が生成されるようになる。そしてこのように生成された感情表現がなされた無意味語の発話文をロボット装置が発話することにより、人間は、ロボット装置の発話内容自体はわからないが、ロボット装置の感情を知ることができるようになる。そして、そのような発話が、さらに毎回異なるものとなるので、人間は、ロボット装置との対話を常に新鮮に感じることができる。次に、本発明の実施の形態であるロボット装置について説明し、その後、そのようなロボット装置への上述の発話のアルゴリズムの実装形態を具体的に説明する。

【0053】なお、実施の形態では、感情に応じたパラメータの制御を、感情に対応して予め用意しているパラメータからなるテーブルを実際の感情により切り換えることにより実現しているが、感情に応じたパラメータの制御がこの実施の形態に限定されないことはいふまでもない。

【0054】(3) 本実施の形態によるロボット装置の具体例

(3-1) ロボット装置の構成

以下、本発明のより具体的な実施の形態として、4本足の自律型ペットロボットに本発明を適用した例について、図面を参照しながら詳細に説明する。このペット型ロボット装置のソフトウェアに感情・本能モデルを導入し、より生物に近い行動を得ることかできるようにしている。本実施の形態では実際に動作をするロボットを用いているが、無意味語による発話はスピーカを持つコンピュータ・システムであれば容易に実現可能であり、人間と機械とのインタラクション（或いは対話）の場で有効な機能である。従って本発明の適用範囲はロボットシステムに限られるものではない。

【0055】具体例としてのロボット装置は、図6に示すように、「犬」を模した形状のいわゆるペットロボットとされ、胴体部ユニット2の前後左右にそれぞれ脚部ユニット3A、3B、3C、3Dが連結されると共に、胴体部ユニット2の前端部及び後端部にそれぞれ頭部ユニット4及び尻尾部ユニット5が連結されて構成されている。

【0056】胴体部ユニット2には、図7に示すように、CPU (Central Processing Unit) 10、DRAM (Dynamic Random Access Memory) 11、フラッシュROM (Read Only Memory) 12、PC (Personal Computer) カードインターフェース回路13及び信号処理回路14が内部バス15を介して相互に接続されることにより形成されたコントロール部16と、このロボット装置1の動力源としてのバッテリー17とが収納されている。また、胴体部ユニット2には、ロボット装置1の向きや動きの加速度を検出するための角速度センサ18及び加速度センサ19なども収納されている。

【0057】また、頭部ユニット4には、外部の状況を撮像するためのCCD (Charge Coupled Device) カメラ20と、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ21と、前方に位置する物体までの距離を測定するための距離センサ22と、外部音を集音するためのマイクロホン23と、鳴き声等の音声出力するためのスピーカ24と、ロボット装置1の「目」に相当するLED (Light Emitting Diode) (図示せず) などがそれぞれ所定位置に配置されている。

【0058】さらに、各脚部ユニット3A～3Dの関節部分や各脚部ユニット3A～3D及び胴体部ユニット2の各連結部分、頭部ユニット4及び胴体部ユニット2の連結部分、並びに尻尾部ユニット5の尻尾5Aの連結部分などにはそれぞれ自由度数分のアクチュエータ25₁～25_n及びポテンシオメータ26₁～26_nが配設されている。例えば、アクチュエータ25₁～25_nはサーボモータを構成として有している。サーボモータの駆動により、脚部ユニット3A～3Dが制御されて、目標の姿勢或いは動作に遷移する。

【0059】そして、これら角速度センサ18、加速度センサ19、タッチセンサ21、距離センサ22、マイクロホン23、スピーカ24及び各ポテンシオメータ26₁～26_nなどの各種センサ並びにLED及び各アクチュエータ25₁～25_nは、それぞれ対応するハブ27₁～27_nを介してコントロール部16の信号処理回路14と接続され、CCDカメラ20及びバッテリー17は、それぞれ信号処理回路14と直接接続されている。

【0060】信号処理回路14は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス15を介して

DRAM11内の所定位置に順次格納する。また信号処理回路14は、これと共にバッテリー17から供給されるバッテリー残量を表すバッテリー残量データを順次取り込み、これをDRAM11内の所定位置に格納する。

【0061】このようにしてDRAM11に格納された各センサデータ、画像データ、音声データ及びバッテリー残量データは、この後CPU10がこのロボット装置1の動作制御を行う際に利用される。

【0062】實際上CPU10は、ロボット装置1の電源が投入された初期時、胴体部ユニット2の図示しないPCカードスロットに装填されたメモリカード28又はフラッシュROM12に格納された制御プログラムをPCカードインターフェース回路13を介して又は直接読み出し、これをDRAM11に格納する。

【0063】また、CPU10は、この後上述のように信号処理回路14よりDRAM11に順次格納される各センサデータ、画像データ、音声データ及びバッテリー残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。

【0064】さらに、CPU10は、この判断結果及びDRAM11に格納した制御プログラムに基づいて続く行動を決定すると共に、当該決定結果に基づいて必要なアクチュエータ25₁～25_nを駆動させることにより、頭部ユニット4を上下左右に振らせたり、尻尾部ユニット5の尻尾5Aを動かしたり、各脚部ユニット3A～3Dを駆動させて歩行させるなどの行動を行わせる。

【0065】また、この際CPU10は、必要に応じて音声データを生成し、これを信号処理回路14を介して音声信号としてスピーカ24に与えることにより当該音声信号に基づく音声を外部に出力させたり、上述のLEDを点灯、消灯又は点滅させる。

【0066】このようにしてこのロボット装置1においては、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動し得るようになされている。

【0067】(3-2) 制御プログラムのソフトウェア構成

ここで、ロボット装置1における上述の制御プログラムのソフトウェア構成は、図8に示すようになる。この図8において、デバイス・ドライバ・レイヤ30は、この制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット31から構成されている。この場合、各デバイス・ドライバは、CCDカメラ20(図7)やタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

【0068】また、ロボティック・サーバ・オブジェクト32は、デバイス・ドライバ・レイヤ30の最下位層に位置し、例えば上述の各種センサやアクチュエータ2

5₁～25_n等のハードウェアにアクセスするためのインターフェースを提供するソフトウェア群でなるバーチャル・ロボット33と、電源の切換えなどを管理するソフトウェア群でなるパワーマネージャ34と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ35と、ロボット装置1の機構を管理するソフトウェア群でなるデザインド・ロボット36とから構成されている。

【0069】マネージャ・オブジェクト37は、オブジェクト・マネージャ38及びサービス・マネージャ39から構成されている。オブジェクト・マネージャ38は、ロボティック・サーバ・オブジェクト32、ミドル・ウェア・レイヤ40、及びアプリケーション・レイヤ41に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ39は、メモリカード28(図7)に格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【0070】ミドル・ウェア・レイヤ40は、ロボティック・サーバ・オブジェクト32の上位層に位置し、画像処理や音声処理などのこのロボット装置1の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ41は、ミドル・ウェア・レイヤ40の上位層に位置し、当該ミドル・ウェア・レイヤ40を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置1の行動を決定するためのソフトウェア群から構成されている。

【0071】なお、ミドル・ウェア・レイヤ40及びアプリケーション・レイヤ41の具体的なソフトウェア構成をそれぞれ図9に示す。

【0072】ミドル・ウェア・レイヤ40は、図9に示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール50～58並びに入力セマンティクスコンバータモジュール59などを有する認識系60と、出力セマンティクスコンバータモジュール68並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、LED点灯用及び音再生用の各信号処理モジュール61～67などを有する出力系69とから構成されている。

【0073】認識系60の各信号処理モジュール50～58は、ロボティック・サーバ・オブジェクト32のバーチャル・ロボット33によりDRAM11(図7)から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール59に与える。ここで、例えば、バーチャル・ロボット33は、所定の通信規約によって、信号の授受或いは変換をする部分として構成さ

れている。

【0074】入力セマンティクスコンバータモジュール59は、これら各信号処理モジュール50～58から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」などの自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ41（図7）に出力する。

【0075】アプリケーション・レイヤ41は、図10に示すように、行動モデルライブラリ70、行動切換えモジュール71、学習モジュール72、感情モデル73及び本能モデル74の5つのモジュールから構成されている。ここで、感情モデル73が、外部からの刺激等により状態が変化される感情の状態を変化させるモデルであって、このような感情モデルにより決定される感情に応じて上述したような発話文への感情表現の重畳がなされる。また、このような感情モデル73や本能モデル74等の状態の監視、すなわちその状態の判別等は、CPU10等の制御手段によってなされる。

【0076】行動モデルライブラリ70には、図11に示すように、「バッテリー残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」などの予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデル70₁～70_nが設けられている。

【0077】そして、これら行動モデル70₁～70_nは、それぞれ入力セマンティクスコンバータモジュール59から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したときなどに、必要に応じて後述のように感情モデル73に保持されている対応する情動のパラメータ値や、本能モデル74に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換えモジュール71に出力する。

【0078】なお、この実施の形態の場合、各行動モデル70₁～70_nは、次の行動を決定する手法として、図12に示すような1つのノード（状態）NODE₀～NODE_nから他のどのノードNODE₀～NODE_nに遷移するかを各ノードNODE₀～NODE_nに間を接続するアークARC₁～ARC_nに対してそれぞれ設定された遷移確率P₁～P_nに基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【0079】具体的に、各行動モデル70₁～70_nは、それぞれ自己の行動モデル70₁～70_nを形成するノードNODE₀～NODE_nにそれぞれ対応させて、これらノードNODE₀～NODE_nごとに図13

に示すような状態遷移表80を有している。

【0080】この状態遷移表80では、そのノードNODE₀～NODE_nにおいて遷移条件とする入力イベント（認識結果）が「入力イベント名」の行に優先順に列記され、その遷移条件についてのさらなる条件が「データ名」及び「データ範囲」の行における対応する列に記述されている。

【0081】したがって、図13の状態遷移表80で表されるノードNODE₁₀₀では、「ボールを検出（BALL）」という認識結果が与えられた場合に、当該認識結果と共に与えられるそのボールの「大きさ（SIZE）」が「0から1000」の範囲であることや、「障害物を検出（OBSTACLE）」という認識結果が与えられた場合に、当該認識結果と共に与えられるその障害物までの「距離（DISTANCE）」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

【0082】また、このノードNODE₁₀₀では、認識結果の入力がない場合においても、行動モデル70₁～70_nが周期的に参照する感情モデル73及び本能モデル74にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル73に保持された「喜び（JOY）」、「驚き（SURPRISE）」若しくは「悲しみ（SADNESS）」のいずれかのパラメータ値が「50から100」の範囲であるときには他のノードに遷移することができるようになっている。

【0083】また、状態遷移表80では、「他のノードへの遷移確率」の欄における「遷移先ノード」の列にそのノードNODE₀～NODE_nから遷移できるノード名が列記されていると共に、「入力イベント名」、「データ値」及び「データの範囲」の行に記述された全ての条件が揃ったときに遷移できる他の各ノードNODE₀～NODE_nへの遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノードNODE₀～NODE_nに遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は100

[%]となっている。

【0084】したがって、図13の状態遷移表80で表されるノードNODE₁₀₀では、例えば「ボールを検出（BALL）」し、そのボールの「SIZE（大きさ）」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30 [%]」の確率で「ノードNODE₁₂₀（node 120）」に遷移でき、そのとき「ACTION1」の行動が出力されることとなる。

【0085】各行動モデル70₁～70_nは、それぞれこのような状態遷移表80として記述されたノードNODE₀～NODE_nがいくつも繋がるようにして構成されており、入力セマンティクスコンバータモジュール

59から認識結果が与えられたときなどに、対応するノード $NODE_0 \sim NODE_n$ の状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換えモジュール 71 に出力するようになされている。

【0086】図 10 に示す行動切換えモジュール 71 は、行動モデルライブラリ 70 の各行動モデル $70_1 \sim 70_n$ からそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデル $70_1 \sim 70_n$ から出力された行動を選択し、当該行動を実行すべき旨のコマンド（以下、これを行動コマンドという。）をミドル・ウェア・レイヤ 40 の出力セマンティクスコンバータモジュール 68 に送出する。なお、この実施の形態においては、図 11 において下側に表記された行動モデル $70_1 \sim 70_n$ ほど優先順位が高く設定されている。

【0087】また、行動切換えモジュール 71 は、行動完了後に出力セマンティクスコンバータモジュール 68 から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール 72、感情モデル 73 及び本能モデル 74 に通知する。

【0088】一方、学習モジュール 72 は、入力セマンティクスコンバータモジュール 59 から与えられる認識結果のうち、「叩かれた」や「撫でられた」など、使用者からの働きかけとして受けた教示の認識結果を入力する。

【0089】そして、学習モジュール 72 は、この認識結果及び行動切換えモジュール 71 からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときに

はその行動の発現確率を上昇させるように、行動モデルライブラリ 70 における対応する行動モデル $70_1 \sim 70_n$ の対応する遷移確率を変更する。

【0090】他方、感情モデル 73 は、「喜び (joy)」、「悲しみ (sadness)」、「怒り (anger)」、「驚き (surprise)」、「嫌悪 (disgust)」及び「恐れ (fear)」の合計 6 つの情動について、各情動ごとにその情動の強さを表すパラメータを保持している。そして、感情モデル 73 は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール 59 から与えられる「叩かれた」及び「撫でられた」などの特定の認識結果と、経過時間及び行動切換えモジュール 71 からの通知などに基づいて周期的に更新する。

【0091】具体的には、感情モデル 73 は、入力セマンティクスコンバータモジュール 59 から与えられる認識結果と、そのときのロボット装置 1 の行動と、前回更新してからの経過時間などに基づいて所定の演算式により算出されるそのときのその情動の変動量を ΔE

$[t]$ 、現在のその情動のパラメータ値を $E[t]$ 、その情動の感度を表す係数を k_e として、(1) 式によって次の周期におけるその情動のパラメータ値 $E[t+1]$ を算出し、これを現在のその情動のパラメータ値 $E[t]$ と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル 73 は、これと同様にして全ての情動のパラメータ値を更新する。

【0092】

【数 1】

$$E[t+1] = E[t] + k_e \times \Delta E[t]$$

... (1)

【0093】なお、各認識結果や出力セマンティクスコンバータモジュール 68 からの通知が各情動のパラメータ値の変動量 $\Delta E[t]$ にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量 ΔE

$[t]$ に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量 ΔE

$[t]$ に大きな影響を与えるようになっている。

【0094】ここで、出力セマンティクスコンバータモジュール 68 からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル 73 は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール 68 からの通知は、上述した学習モジュール 72 にも入力されており、学習モジュール 72 は、その通知に基づいて行動モデル $70_1 \sim 70_n$ の対応する遷移確率を変更する。

【0095】なお、行動結果のフィードバックは、行動

切換えモジュレータ 71 の出力（感情が付加された行動）によりなされるものであってもよい。

【0096】一方、本能モデル 74 は、「運動欲 (exercise)」、「愛情欲 (affection)」、「食欲 (appetite)」及び「好奇心 (curiosity)」の互いに独立した 4 つの欲求について、これら欲求ごとにその欲求の強さを表すパラメータを保持している。そして、本能モデル 74 は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール 59 から与えられる認識結果や、経過時間及び行動切換えモジュール 71 からの通知などに基づいて周期的に更新する。

【0097】具体的には、本能モデル 74 は、「運動欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール 68 からの通知などに基づいて所定の演算式により算出されるそのときのその欲求の変動量を $\Delta I[k]$ 、現在のその欲求のパラメータ値を $I[k]$ 、その欲求の感度を表す係数 k_i として、所定期期で (2) 式を用いて次の周期におけるその欲求のパラメータ値 $I[k+1]$ を算出し、この演算結果を現在のその欲求のパラメ

ータ値 $I[k]$ と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル74は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

【0098】

【数2】

$$I[k+1] = I[k] + k_i \times \Delta I[k] \quad \dots (2)$$

【0099】なお、認識結果及び出力セマンティクスコンバータモジュール68からの通知などが各欲求のパラメータ値の変動量 $\Delta I[k]$ にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール68からの通知は、「疲れ」のパラメータ値の変動量 $\Delta I[k]$ に大きな影響を与えるようになっている。

【0100】なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数 k_e 、 k_i の値も各情動及び各欲求ごとに個別に設定されている。

【0101】一方、ミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68は、図9に示すように、上述のようにしてアプリケーション・レイヤ41の行動切換えモジュール71から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系69の対応する信号処理モジュール61～67に与える。

【0102】そしてこれら信号処理モジュール61～67は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動を行うために対応するアクチュエータ25₁～25_n（図7）に与えるべきサーボ指令値や、スピーカ24（図7）から出力する音の音声データ及び又は「目」のLEDに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト32のバーチャル・ロボット33及び信号処理回路14（図7）を順次介して対応するアクチュエータ25₁～25_n又はスピーカ24又はLEDに順次送出する。

【0103】このようにしてロボット装置1においては、制御プログラムに基づいて、自己（内部）及び周囲（外部）の状況や、使用者からの指示及び働きかけに応じた自律的な行動を行うことができるようになされている。

【0104】（3-3）ロボット装置への発話のアルゴリズムの実装

上述したようにロボット装置を構成することができる。上述の発話のアルゴリズムは、このようなロボット装置1の図9中の音再生モジュール67として実装される。

【0105】音再生モジュール67では、上位の部分（例えば、行動モデル）にて決定された音出力コマンド

（例えば、「喜びで発話せよ」など）を受け、実際の音声時系列データを生成し、順にバーチャルロボット33のスピーカデバイスに対してデータを送信する。これによりロボット装置において、図7に示すスピーカ24から感情表現がなされた無意味語からなる発話文が発せられる。

【0106】感情に合わせた発話コマンドを生成する行動モデル（以下、発話行動モデルという。）について説明する。発話行動モデルは、図10に示した行動モデルライブラリ70にて一の行動モデルとして用意されている。

【0107】発話行動モデルでは、感情モデル73や本能モデル74から常に最新のパラメータ値を参照して、そのような各パラメータ値に基づいて図10に示すような状態遷移表80を利用して、発話内容を決定している。すなわち、ある状態からの遷移条件として感情の値を用い、その感情に即した発話行動を遷移時に実行するようにしている。

【0108】発話行動モデルが使用する状態遷移表は、例えば、図14に示すように表現することができる。なお、図14に示す発話行動モデルに使用する状態遷移表は、上述の図13に示した状態遷移表80の表記形成が異なっているが、実質的には異なるものではない。図14のように示される状態遷移表について説明する。

【0109】この例では、ノードが“node XXX”から他のノードへの遷移条件として喜び（HAPPY）、悲しみ（SAD）、怒り（ANGER）及びタイムアウト（TIMEOUT）が与えられている。そして、喜び（HAPPY）、悲しみ（SAD）、怒り（ANGER）及びタイムアウト（TIMEOUT）への遷移条件としての具体的な数値が、それぞれ $HAPPY > 70$ 、 $SAD > 70$ 、 $ANGER > 70$ 及び $TIMEOUT = timeout.1$ として与えられている。ここで、 $timeout.1$ は数値であり、例えば所定時間を示す値である。

【0110】また、ノードが“node XXX”から遷移可能なノードとして、node YYY、node ZZZ、node WWW、node VVVが用意されており、そのような各ノードに対して実行される行動がそれぞれ「バンザイ（BANZAI）」、「落ち込む（OTIKOMU）」、「ぶるぶる（BURUBURU）」及び「あくび（AKUBI）」として割り当てられている。

【0111】ここで、「万歳（BANZAI）」の表現行動は、「喜び」が感情表現される発話（talk_happy）をし、また前脚等による万歳の動作（motion_banzai）をし、さらに尻尾を振る動作（motion_swingtail）をするものとして定義している。ここで「喜び」の感情表現をした発話をするために、上述したような予め用意されている喜びの感情表現のためのパラメータを使用する。すなわち、先に説明している発話のアルゴリズムに基づいて喜びの発話を行う。

【0112】また、「落ち込む（OTIKOMU）」の表現行

動は、「悲しみ」が感情表現される発話 (talk_sad) をし、またいわゆるいじけた動作 (motion_ljijji) をするものとして定義している。ここで「悲しみ」の感情表現をした発話をするために、上述したような予め用意されている悲しみの感情表現のためのパラメータを使用する。すなわち、先に説明している発話のアルゴリズムに基づいて悲しみの発話を行う。

【0113】また、「ぶるぶる (BURUBURU)」の表現行動は、「怒り」が感情表現される発話 (talk_anger) をし、また怒りのために震えている動作 (motion_buruburu) をするものとして定義している。ここで「怒り」の感情表現をした発話をするために、上述したような予め用意されている怒りの感情表現のためのパラメータを使用する。すなわち、先に説明している発話のアルゴリズムに基づいて怒りの発話を行う。

【0114】また、「あくび (AKUBI)」の表現行動は、何もなく退屈なのであくびをする動作 (motion_aku bi) として定義している。

【0115】このように遷移可能な各ノードにおいて実行される各行動が定義されており、そのような各ノードへの遷移については、確率テーブルによって決定されている。すなわち、遷移条件に合致した場合の行動確率が記述された確率テーブルにて、各ノードへの遷移を決定している。

【0116】図14に示す例では、喜び (HAPPY) の場合、すなわちHAPPYの値が所定の閾値とされる70を超えた場合には100%の確率で「万歳 (BANZAI)」の表現行動で選択される。また、悲しみ (SAD) の場合、すなわちSADの値が所定の閾値とされる70を超えた場合には、100%の確率で「落ち込む (OTIKOMU)」の表現行動で選択される。また、怒り (ANGER) の場合、すなわちANGERの値が所定の閾値とされる70を超えた場合には、100%の確率で「ぶるぶる (BURUBURU)」の表現行動が選択される。そして、タイムアウト (TIMEOUT) の場合、すなわちTIMEOUTの値が所定の閾値とされるtimout.1となった場合には、100%の確率で「あくび (AKUBI)」の表現行動が選択される。なお、本例では、全て100%の確率で行動を選択される場合を示しており、すなわち必ず行動が発言される場合を例にしているが、これに限定されることはない。すなわち例えば、喜び (HAPPY) の場合に、「万歳 (BANZAI)」の行動を70%で選択するようにしてもよい。

【0117】以上のように発話行動モデルの状態遷移表を定義することにより、ロボット装置に感情に見合った発話をその他のセンサの入力やロボットの状態に合わせて自由に制御することかできるようになる。

【0118】なお、上述の実施の形態では、感情によって制御されるパラメータとして、持続時間、ピッチ及び音量を例に挙げて説明した。しかし、これに限定されることなく、感情によって影響される文章構成因子をパラ

メータとして用いることもできる。

【0119】また、上述の実施の形態の説明では、ロボット装置の感情モデルが喜び、怒り等の感情によって構成される場合を説明した。しかし、このような例に挙げた感情によって感情モデルが構成されることに限定されるものではなく、感情に影響を与える他の因子によって構成することもできる。そして、この場合、このような他の因子によって、文章を構成するパラメータの制御を行う。

【0120】

【発明の効果】本発明に係る音声合成方法は、発音主体の感情モデルの感情状態を判別する感情判別工程と、音声として発する内容を表す発話文を出力する発話文出力工程と、感情判別工程により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御工程と、発話文出力工程により出力された発話文を音声合成部に入力して制御されたパラメータに基づいて音声合成する音声合成工程とを有することにより、発音主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生成することができる。

【0121】また、本発明に係る音声合成装置は、発音主体の感情モデルの感情状態を判別する感情判別手段と、音声として発する内容を表す発話文を出力する発話文出力手段と、感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成する音声合成手段とを備えることにより、発音主体の感情モデルの感情状態を判別する感情判別手段により判別された感情状態に応じて音声合成のためのパラメータをパラメータ制御手段により制御して、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成手段により音声合成することができる。これにより、音声合成装置は、発音主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生成することができる。

【0122】また、本発明に係るロボット装置は、動作に起因する感情モデルと、感情モデルの感情状態を判別する感情判別手段と、音声として発する内容を表す発話文を出力する発話文出力手段と、感情判別手段により判別された感情状態に応じて音声合成のためのパラメータを制御するパラメータ制御手段と、発話文出力手段により出力された発話文が供給され、制御されたパラメータに基づいて音声合成する音声合成手段とを備えることにより、動作に起因する感情モデルの感情状態を判別する感情判別手段により判別された感情状態に応じて音声合成のためのパラメータをパラメータ制御手段により制御して、発話文出力手段により出力された発話文が供給さ

れ、制御されたパラメータに基づいて音声合成手段により音声合成することができる。これにより、ロボット装置は、発声主体の感情モデルの感情状態に応じて制御した音声合成のためのパラメータに基づいて発音主体の発話文を生成することができる。

【図面の簡単な説明】

【図 1】本発明に係る音声合成方法の実施の形態の基本構成を示すフローチャートである。

【図 2】各音素の持続時間とピッチとの関係を示す図である。

【図 3】音声合成による発話文の作成のためのプログラムであって、その前半部分を示す図である。

【図 4】音声合成による発話文の作成のためのプログラムであって、その後半部分を示す図である。

【図 5】特徴空間或いは作用平面における各感情のクラスの関係を示す図である。

【図 6】本発明の実施の形態であるロボット装置の外観構成を示す斜視図である。

【図 7】上述のロボット装置の回路構成を示すブロック図である。

【図 8】上述のロボット装置のソフトウェア構成を示す

ブロック図である。

【図 9】上述のロボット装置のソフトウェア構成におけるミドル・ウェア・レイヤの構成を示すブロック図である。

【図 10】上述のロボット装置のソフトウェア構成におけるアプリケーション・レイヤの構成を示すブロック図である。

【図 11】上述のアプリケーション・レイヤの行動モデルライブラリの構成を示すブロック図である。

【図 12】ロボット装置の行動決定のための情報となる有限確率オートマトンを説明するために使用した図である。

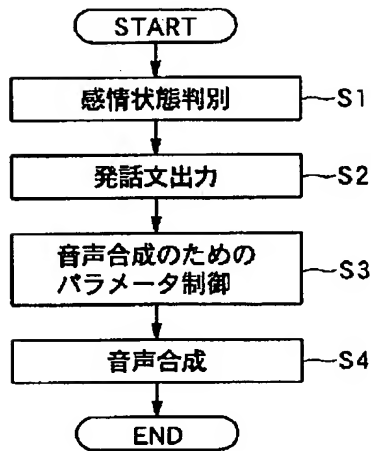
【図 13】有限確率オートマトンの各ノードに用意された状態遷移表を示す図である。

【図 14】発話行動モデルの状態遷移表を示す図である。

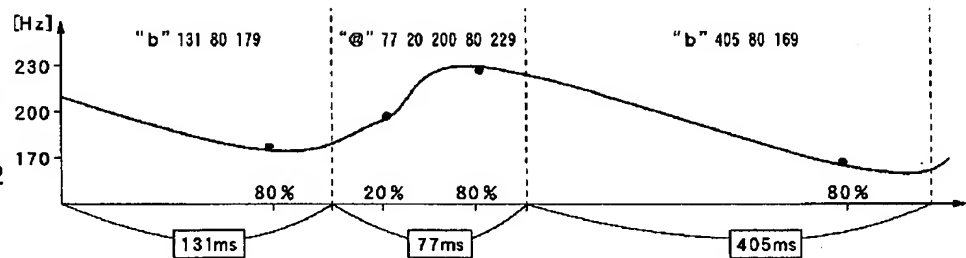
【符号の説明】

1 ロボット装置、10 CPU、14 信号処理回路、24 スピーカ、70 行動モデル、73 感情モデル

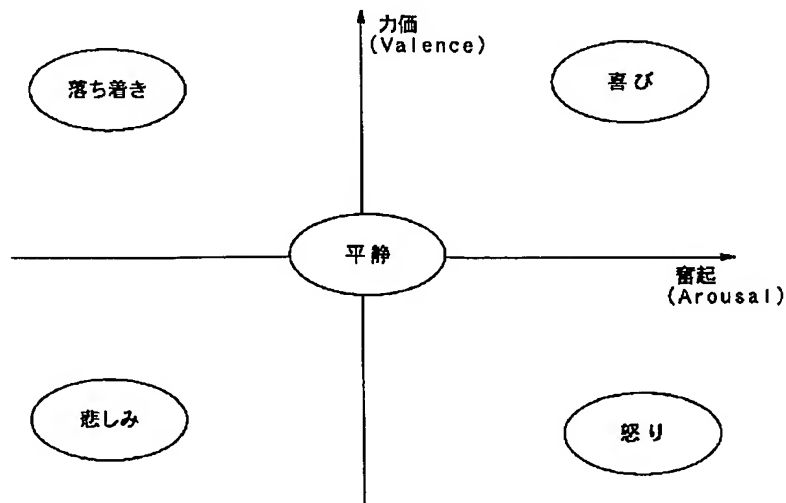
【図 1】



【図 2】



【図 5】



【図3】

```

1 Choose the number of words of the sentence (random number between 2 and MAXWORDS);
2 Create the words :
3   For each word, choose the number of syllables
4     (random number between 2 and MAXSYLL), and
5     decides with probability PROBACCENT whether the word is accented or not ;
6 If the word is accented then choose randomly one
7   of its syllables and mark it as accented ;
8 Create the syllables :
9   For each syllable
10    choose whether this is a CV or a CCV syllable
11      (CV syllable have probability 0.8) ;
12    instantiate the C's and V by picking randomly a
13      consonant or vowel in the phoneme database ;
14    set the duration of each phoneme to MEANDUR + random(DURVAR) ;
15    let e = MEANPITCH + random(PITCHVAR)
16      set the pitch of consonants to e - PITCHVAR
17      set the pitch of vowels to e + PITCHVAR
18    if the syllable is accented then
19      add DURVAR to the duration of its phonemes ;
20      if DEFAULTCONTOUR = rising
21        set the pitch of consonants to MAXPITCH - PITCHVAR
22        set the pitch of the vowel to MAXPITCH + PITCHVAR
23      if DEFAULTCONTOUR = falling
24        set the pitch of consonants to MAXPITCH + PITCHVAR
25        set the pitch of the vowel to MAXPITCH - PITCHVAR
26      if DEFAULTCONTOUR = stable
27        set the pitch of phonemes to MAXPITCH
28

```

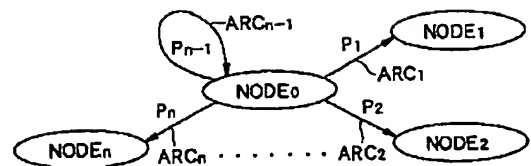
【図4】

```

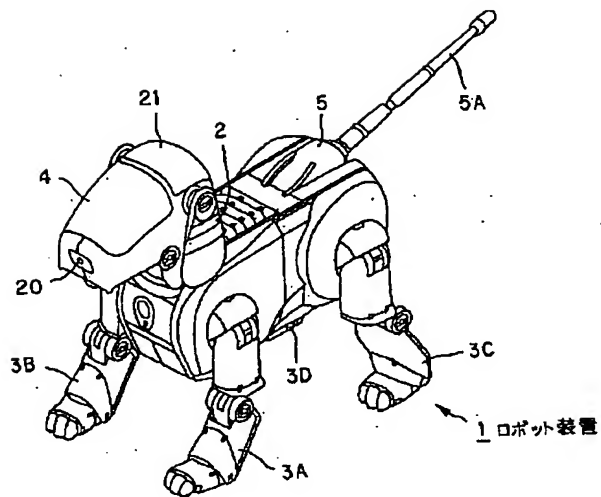
29 Change the contour of the last word :
30 if not LASTWORDACCENTED
31   let e = PITCHVAR/2
32   if CONTOURLASTWORD = FALLING
33     for each syllable in word
34       add -(i+1)*e pitch of phonemes to their value
35         (i = index of phoneme in syllable)
36       e = e + e
37   if CONTOURLASTWORD = RISING
38     for each syllable in word
39       add +(i+1)*e pitch of phonemes to their value
40         (i = index of phoneme in syllable)
41   e = e + e
42   else
43     if CONTOURLASTWORD = FALLING
44       for each syllable in word
45         add DURVAR to the duration of its phonemes ;
46         set the pitch of consonants to MAXPITCH + PITCHVAR
47         set the pitch of the vowel to MAXPITCH - PITCHVAR
48     if CONTOURLASTWORD = RISING
49       for each syllable in word
50         add DURVAR to the duration of its phonemes ;
51         set the pitch of consonants to MAXPITCH - PITCHVAR
52         set the pitch of the vowel to MAXPITCH + PITCHVAR
53 Set the loudness volume of the complete sentence to VOLUME.

```

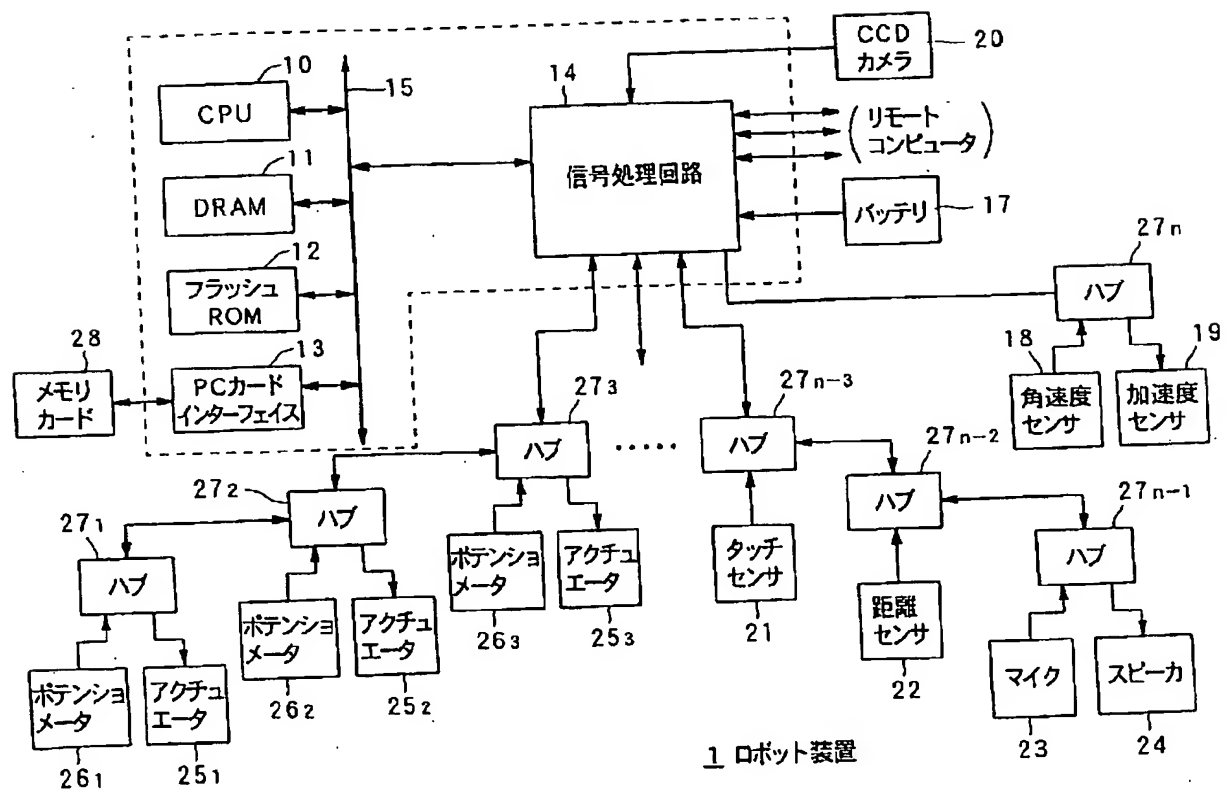
【図12】



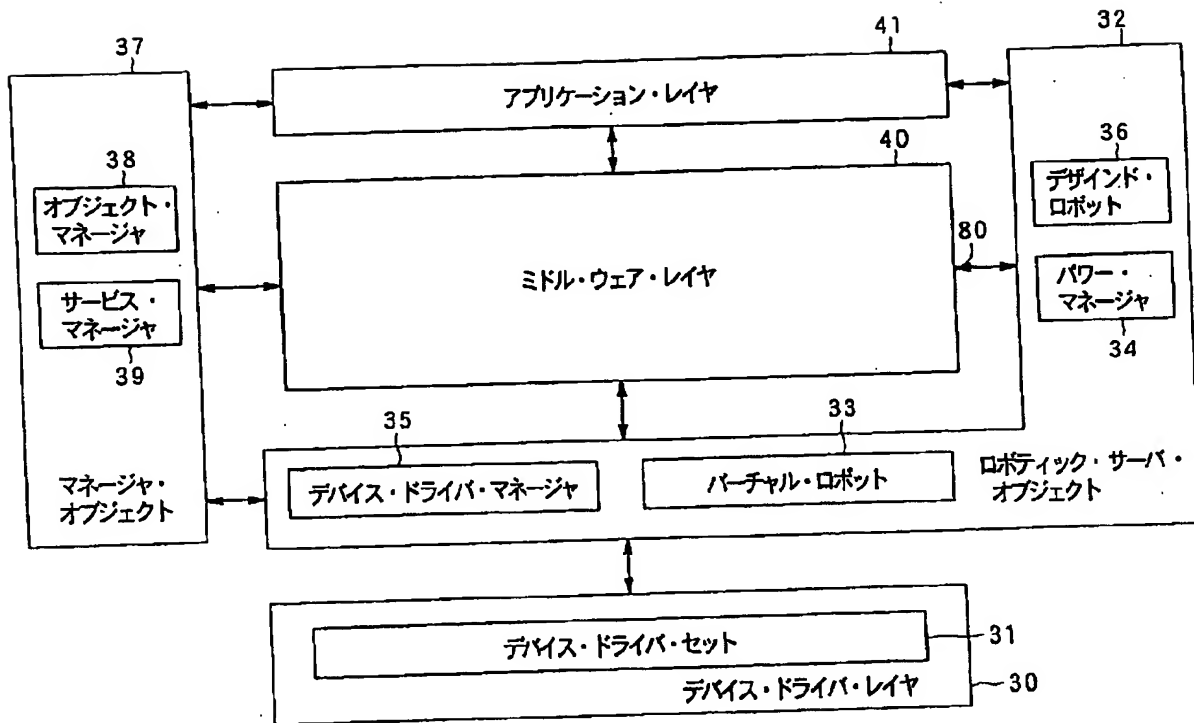
【図6】



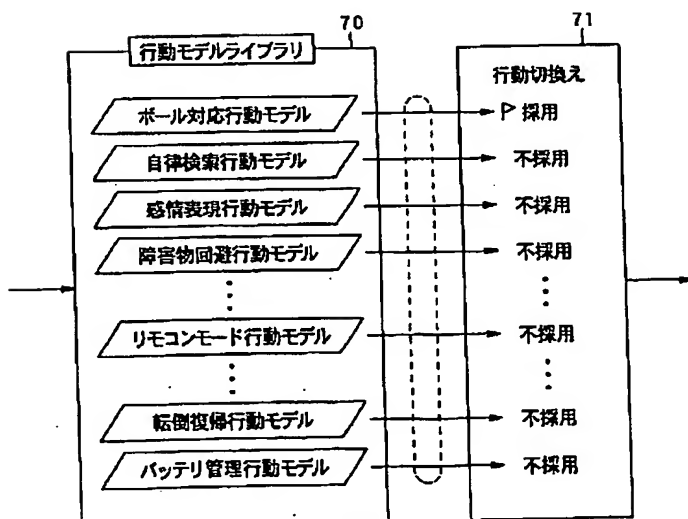
【図7】



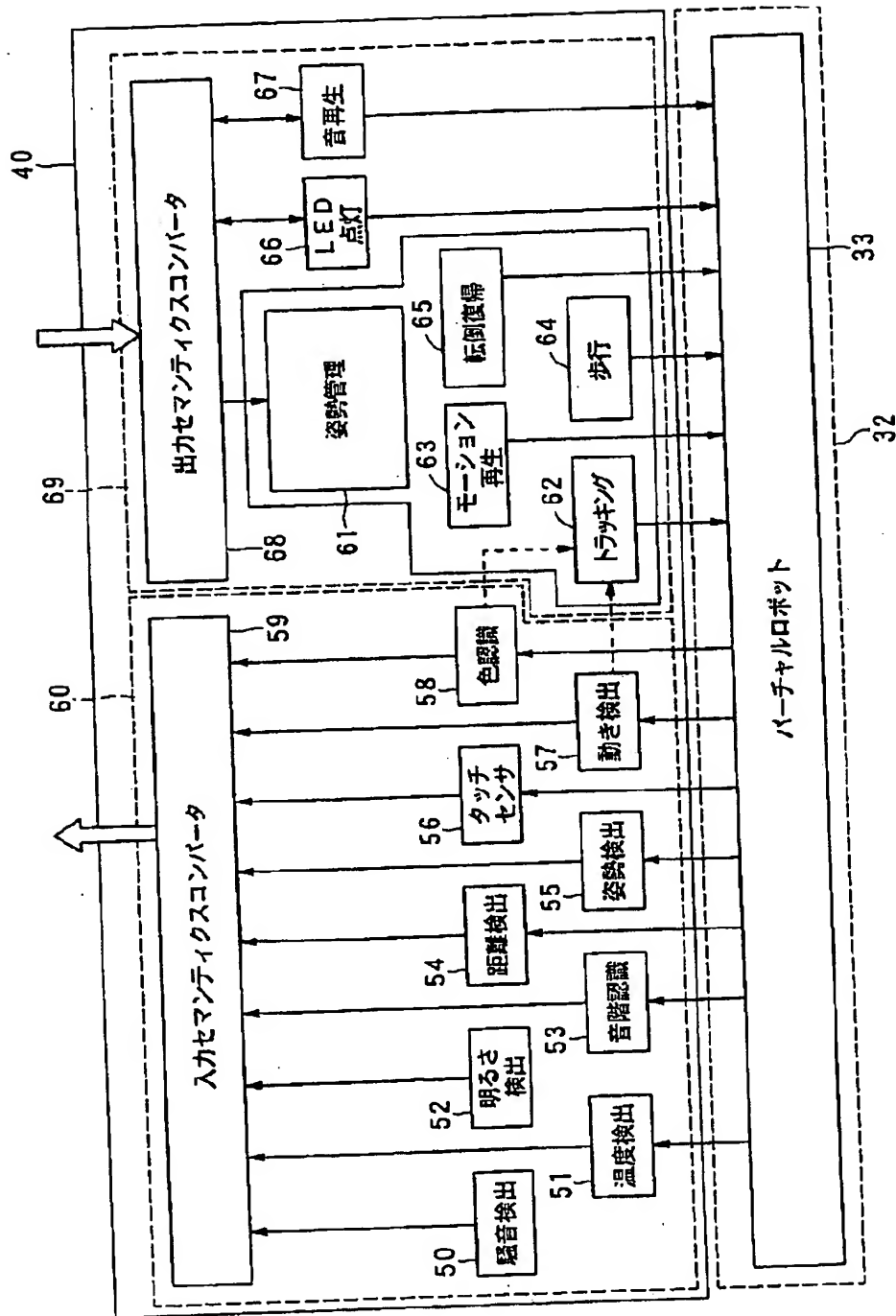
【図8】



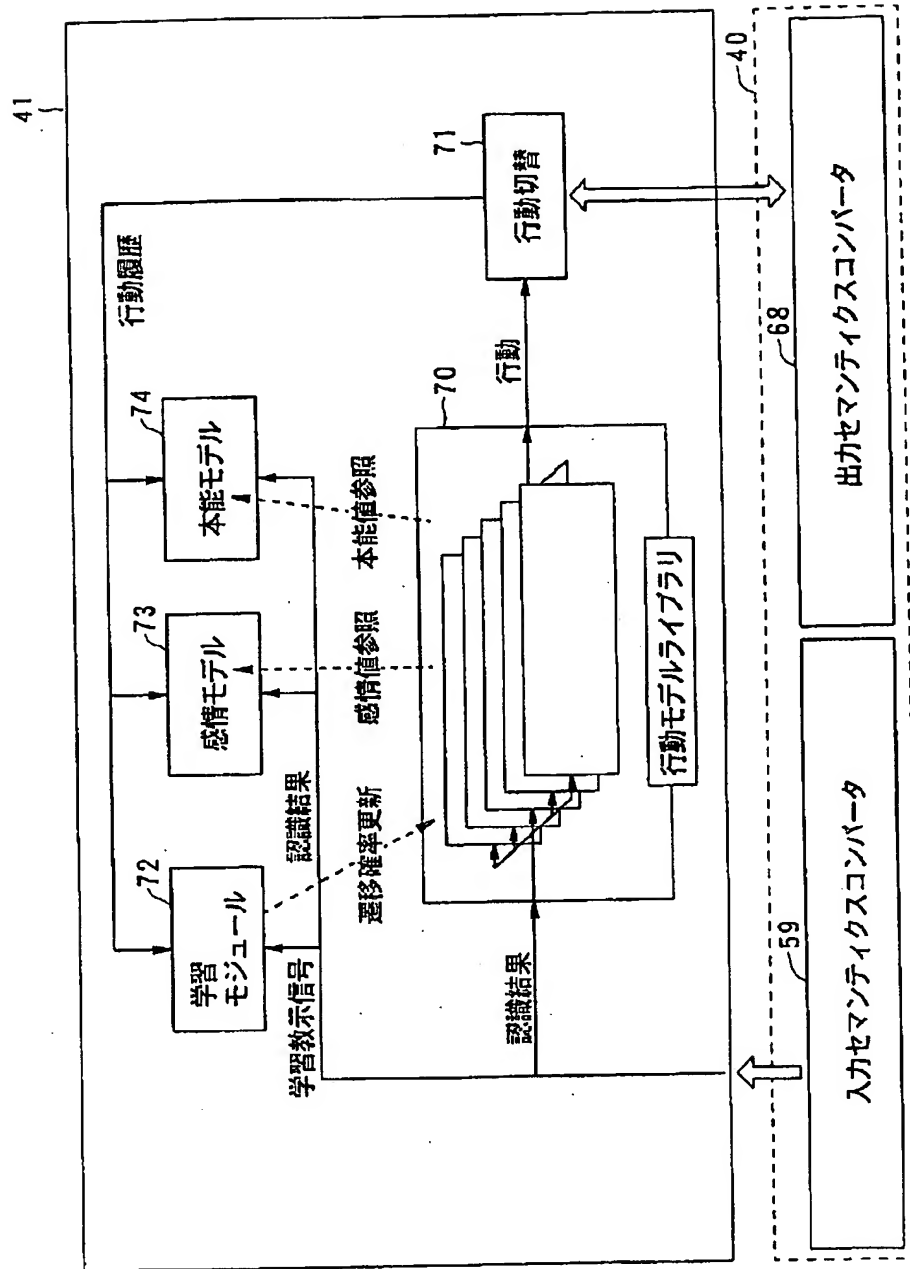
【図11】



【図9】



【図10】



【図13】

	入カイベント名	データ名	データの範囲	他のノードへの遷移確率 DI				n
node 100				A	B	C	D	node 600
遷移先ノード				node 120	node 120	node 1000		
出力行動				ACTION 1	ACTION 2	MOVE BACK		ACTION 4
1	BALL	SIZE	0.1000	30%				
2	PAT				40%			
3	HIT				20%			
4	MOTION					50%		
5	OBSTACLE	DISTANCE	0.100			100%		
6		JOY	50.100					
7		SUPRISE	50.100					
8		SUDNESS	50.100					

80

【図14】

動作の定義等	node XXX				
	Condition-label:				
	HAPPY	happy>70			
	SAD	sad >70			
	ANGER	anger>70			
	TIMEOUT	timeout.1			
	Arc-label:				
	BAZAI	node YYY	talk_happy, motion_banzai, motion_swingtail		
	OTIKOMU	node ZZZ	talk_sad, motion_ljiji		
	BURUBURU	node WWW	talk_anger, motion_buruburu		
確率テーブル	AKUBI	node VVV	motion_akubi		
	Probability-table:				
		BANZAI	OTIKOMU	BURUBURU	AKUBI
	HAPPY	100			
	SAD		100		
	ANGER			100	
	TIMEOUT				100

フロントページの続き

(51)Int.Cl.7

G 0 6 F 3/16

G 1 0 L 13/08

識別記号

3 4 0

F I

G 1 0 L 3/00

テーマコード(参考)

E

H

Fターム(参考) 2C150 BA11 CA01 CA02 CA04 DF03
 DF04 DF06 DF33 ED42 ED52
 EF16 EF23 EF29 EF33 EF36
 3C007 AS36 KS11 KS15 KS23 KS24
 KS31 KS36 KS39 KT01 LW12
 MT14 WA04 WA14 WB15 WB27
 5D045 AA07 AA20

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-175091

(43)Date of publication of application : 21.06.2002

(51)Int.Cl. G10L 13/00

A63H 11/00

B25J 5/00

B25J 13/00

G06F 3/16

G10L 13/08

(21)Application number : 2000-372091 (71)Applicant : SONY CORP

(22)Date of filing : 06.12.2000 (72)Inventor : SABE KOTARO

OUDEYER PIERRE YVES

(54) SPEECH SYNTHESIS METHOD AND APPARATUS AND ROBOT
APPARATUS

(57)Abstract:

PROBLEM TO BE SOLVED: To make an auditory emotional expression which more closely approximates the emotional expression of a living being, etc.

SOLUTION: The robot apparatus forms utterance sentences by speech synthesis by an emotion discriminating process step (step S1) of discriminating the emotion state of an emotion model, an utterance sentence outputting process step (step S2) for outputting the utterance sentences indicating the contents to be uttered as a speech, a parameter control process step (step S3) of controlling the parameters for speech synthesis, according to the emotion state discriminated by the emotion discriminating process step and a speech synthesizing process step (step S4) of inputting the utterance sentences

outputted by the utterance sentence outputting process step to the speech synthesizing section and performing speech synthesis according to the controlled parameters.

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JP0 and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The feeling distinction process which is the speech synthesis approach which compounds voice based on the information from the pronunciation subject who has a feeling model at least, and distinguishes the feeling condition of the above-mentioned pronunciation subject's above-mentioned feeling model, The utterance sentence output process which outputs the utterance sentence showing the content emitted as voice, and the parameter-control process which controls the parameter for speech synthesis according to the feeling condition distinguished by the above-mentioned feeling distinction process, The speech synthesis approach characterized by having the speech synthesis process which synthesizes voice based on the parameter by which inputted into the speech synthesis section the utterance sentence outputted by the above-mentioned utterance sentence output process, and control was carried

out [above-mentioned].

[Claim 2] The above-mentioned utterance sentence is the speech synthesis approach according to claim 1 characterized by being the sentence of a meaningless content.

[Claim 3] The above-mentioned utterance sentence output process is the speech synthesis approach according to claim 1 characterized by outputting the above-mentioned utterance sentence and supplying the above-mentioned speech synthesis section when the feeling condition of the above-mentioned feeling model exceeds a predetermined threshold.

[Claim 4] The above-mentioned utterance sentence output process is the speech synthesis approach according to claim 1 characterized by outputting the above-mentioned utterance sentence obtained at random for every utterance, and supplying the above-mentioned speech synthesis section.

[Claim 5] It is the speech synthesis approach according to claim 1 which the above-mentioned utterance sentence has two or more phonemes, changes, and is characterized by the above-mentioned parameter containing the persistence time of the above-mentioned phoneme, a pitch, and sound volume.

[Claim 6] The above-mentioned pronunciation subject is the speech-synthesis approach according to claim 1 which is autonomous mold robot equipment which operates based on the supplied input, and is characterized by to have the

feeling model which originates in the above-mentioned actuation as the above-mentioned feeling model, and to have the feeling model change process of opting for the above-mentioned actuation by changing the condition of the above-mentioned feeling model based on the above-mentioned input.

[Claim 7] A feeling distinction means to be the voice synthesizer which compounds voice based on the information from the pronunciation subject who has a feeling model at least, and to distinguish the feeling condition of the above-mentioned pronunciation subject's above-mentioned feeling model, An utterance sentence output means to output the utterance sentence showing the content emitted as voice, and a parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the above-mentioned feeling distinction means, The voice synthesizer characterized by having the speech synthesis means from which the utterance sentence outputted by the above-mentioned utterance sentence output means is supplied, and synthesizes voice based on the parameter by which control was carried out [above-mentioned].

[Claim 8] The above-mentioned utterance sentence is a voice synthesizer according to claim 7 characterized by being the sentence of a meaningless content.

[Claim 9] The above-mentioned utterance sentence output means is a voice

synthesizer according to claim 7 characterized by outputting the above-mentioned utterance sentence and supplying the above-mentioned speech synthesis means when the feeling condition of the above-mentioned feeling model exceeds a predetermined threshold.

[Claim 10] The above-mentioned utterance sentence output means is a voice synthesizer according to claim 7 characterized by outputting the above-mentioned utterance sentence obtained at random for every utterance, and supplying the above-mentioned speech synthesis means.

[Claim 11] It is the voice synthesizer according to claim 7 which the above-mentioned utterance sentence has two or more phonemes, changes, and is characterized by the above-mentioned parameter containing the persistence time of the above-mentioned phoneme, a pitch, and sound volume.

[Claim 12] The above-mentioned pronunciation subject is the voice synthesizer according to claim 7 which is autonomous mold robot equipment which performs actuation according to the supplied input, and is characterized by to have a feeling model change means opt for the above-mentioned actuation by having the feeling model which originates in the above-mentioned actuation as the above-mentioned feeling model, and changing the condition of the above-mentioned feeling model based on the above-mentioned input.

[Claim 13] The feeling model which is robot equipment of the autonomous mold

which operates based on the supplied input, and originates in the above-mentioned actuation, A feeling distinction means to distinguish the feeling condition of the above-mentioned feeling model, and an utterance sentence output means to output the utterance sentence showing the content emitted as voice, A parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the above-mentioned feeling distinction means, Robot equipment characterized by having the speech synthesis means from which the utterance sentence outputted by the above-mentioned utterance sentence output means is supplied, and synthesizes voice based on the parameter by which control was carried out [above-mentioned].

[Claim 14] The above-mentioned utterance sentence is robot equipment according to claim 13 characterized by being the sentence of a meaningless content.

[Claim 15] The above-mentioned utterance sentence output means is robot equipment according to claim 13 characterized by outputting the above-mentioned utterance sentence and supplying the above-mentioned speech synthesis means when the feeling condition of the above-mentioned feeling model exceeds a predetermined threshold.

[Claim 16] The above-mentioned utterance sentence output means is robot

equipment according to claim 13 characterized by outputting the above-mentioned utterance sentence obtained at random for every utterance, and supplying the above-mentioned speech synthesis means.

[Claim 17] It is robot equipment according to claim 13 which the above-mentioned utterance sentence has two or more phonemes, changes, and is characterized by the above-mentioned parameter containing the persistence time of the above-mentioned phoneme, a pitch, and sound volume.

[Claim 18] Robot equipment according to claim 13 characterized by having a feeling model change means to opt for the above-mentioned actuation by changing the condition of the above-mentioned feeling model based on the above-mentioned input.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the speech synthesis approach for generating the voice which an utterance subject outputs and a voice synthesizer, and the robot equipment that outputs voice to a list.

[0002]

[Description of the Prior Art] In recent years, the robot equipment of for example, the pet mold with which the appearance configuration was formed in animals, such as a dog and a cat, by imitating is offered. There are some which operate autonomously according to the condition of the information from the outside or the interior in such robot equipment.

[0003] For the artificial intelligence (A.I. Artificial Intelligence:artificial intelligence) used for such robot equipment to realize intellectual functions, such as inference and decision, artificially, and to also realize functions, such as feeling and instinct, artificially further is tried. Among visual expression means, acoustic sense expression means, etc. as an expression means to the exterior of such artificial intelligence, using voice is mentioned as an example of an acoustic-sense-thing.

[0004] In such robot equipment, the function to complain of own feeling to human beings (owner etc.) using the utterance is effective. This is because the pet dog and the favorite cat could know from the experience what kind of mood it was now and the one element has judged by utterance of a pet, although human being cannot understand directly what pets, such as a actual dog and a cat, have said.

[0005]

[Problem(s) to be Solved by the Invention] By the way, what is performing the acoustic sense feeling expression to the electronic sound (audible tone) as robot equipment currently supplied to the current commercial scene is known. It is that a short high sound expresses joy and the low sound carried out slowly specifically expresses sadness etc. And music is written beforehand, and by human being's subjectivity, these audible tones can be distributed to each feeling class, and are used for playback. Here, a feeling class is a class of the emotion which is glad and is classified as resentment etc. the acoustic sense feeling expression using the audible tone currently made conventionally -- setting -- (i) -- mechanical (ii) -- the power of expression (iii) which repeats the always same expression is suitable, or unknown -- etc. -- the actual condition is that are mentioned as a point that a point differs from a feeling expression of the pet of living things, such as a dog and a cat, greatly, and the further improvement is desired.

[0006] Then, this invention is made in view of the above-mentioned actual condition, and aims at offer of robot equipment in the speech synthesis approach which enables an acoustic sense feeling expression which is near with the feeling expression of a living thing etc. and a voice synthesizer, and a list.

[0007]

[Means for Solving the Problem] The feeling distinction process which

distinguishes the feeling condition of a pronunciation subject's feeling model in order that the speech synthesis approach concerning this invention may solve an above-mentioned technical problem, The utterance sentence output process which outputs the utterance sentence showing the content emitted as voice, and the parameter-control process which controls the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction process, It has the speech synthesis process which synthesizes voice based on the parameter which inputted into the speech synthesis section the utterance sentence outputted by the utterance sentence output process, and was controlled.

[0008] Such a speech synthesis approach generates a pronunciation subject's utterance sentence based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model.

[0009] Moreover, in order that the voice synthesizer concerning this invention may solve an above-mentioned technical problem A feeling distinction means to distinguish the feeling condition of a pronunciation subject's feeling model, and an utterance sentence output means to output the utterance sentence showing the content emitted as voice, It has a parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction means, and the speech synthesis means from which

the utterance sentence outputted by the utterance sentence output means is supplied, and synthesizes voice based on the controlled parameter.

[0010] A voice synthesizer equipped with such a configuration controls the parameter for speech synthesis by the parameter-control means according to the feeling condition distinguished by feeling distinction means distinguish the feeling condition of a pronunciation subject's feeling model, and the utterance sentence outputted by the utterance sentence output means is supplied, and synthesizes voice from it with a speech-synthesis means based on the controlled parameter. Thereby, a voice synthesizer generates a pronunciation subject's utterance sentence based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model.

[0011] Moreover, in order that the robot equipment concerning this invention may solve an above-mentioned technical problem The feeling model resulting from actuation, and a feeling distinction means to distinguish the feeling condition of a feeling model, An utterance sentence output means to output the utterance sentence showing the content emitted as voice, and a parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction means, It has the speech synthesis means from which the utterance sentence outputted

by the utterance sentence output means is supplied, and synthesizes voice based on the controlled parameter.

[0012] Robot equipment equipped with such a configuration controls the parameter for speech synthesis by the parameter-control means according to the feeling condition distinguished by feeling distinction means distinguish the feeling condition of the feeling model resulting from actuation, and the utterance sentence outputted by the utterance sentence output means is supplied, and synthesizes voice with a speech-synthesis means based on the controlled parameter from it. Thereby, robot equipment generates a pronunciation subject's utterance sentence based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model.

[0013]

[Embodiment of the Invention] First, the semantics of adopting the feeling expression with voice as the speech synthesis approach concerning this invention and equipment, and a list as functions, such as robot equipment of a pet mold, in advance of explanation of the gestalt of desirable operation of robot equipment and the feeling expression with suitable voice are explained.

[0014] (1) Adding a feeling expression to utterance as a function in feeling expression robot equipment with voice etc. works very effectively, in order to raise intimate nature with human being. Moreover, a stimulus can be required of

human being by satisfaction and the dissatisfaction guide peg of not only the improvement in sociality but self only being shown. Such a function turns into the function to act effectively in robot equipment with a learning function.

[0015] About whether on the other hand the acoustic feature of the feeling which human being has, and the voice uttered has correlation the report (Fairbanks G. (1940) --) of Fairbanks Recent experimental investigations of vocal pitch in speech and Journal of the Acoustical Society A report of of America, (11):457-466. and Burkhardt and others () [Burkhardt F.and Sendlmeier W.F.,] [Verification of Acoustic Correlates] There are of Emotional Speech using Formant Synthesis, ISCA Workshop on Speech and Emotion, Belfast 2000., etc. Thus, the report by many researchers is known.

[0016] According to these reports, it turns out that utterance has correlation with psychological conditions or some fundamental emotion classes. Moreover, it is reported that it is difficult to find a difference about the specific feeling which is conversely afraid surprise and is [sadness / tedium,] in it etc. It is connected with a certain physical condition about a certain emotion, and expectable physical effect is easily brought about to utterance.

[0017] For example, when a certain man memorizes the resentment, fear, and joy, the sympathetic nervous system calls forth, and a heart rate and blood pressure rise. In opening, it gets dry and, occasionally shaking takes place to

muscles. At such time, utterance becomes quick greatly, and it will have energy strong against a high frequency component. Moreover, when a certain man senses tedium and sadness, the parasympathetic nervous system is stirred up. A heart rate and blood pressure decrease and many saliva is secreted. As a result, utterance becomes what has the late pitch carried out slowly. And since these bodily features is universal, it is considered that the correlation which approaches neither a race nor culture exists between the fundamental emotion and the acoustical property of utterance.

[0018] moreover Enumeration of the word which does not have semantics to Japanese people and Americans Abelin and others reports [the result of having experimented in how many each other feeling it being made speaking by various feeling and being recognized] (). [Abelin A., Allwood J.,] [Cross Linguistic Interpretation] of Emotional Prosody, Workshop on Emotions in Speech, ISCA Workshop on Speech and Emotion, and Belfast 2000., the report (TickleA. --) of Tickle English and Japanese Speaker's Emotion Vocalizations and Recognition : A Comparison HighlightingVowel Quality and SCAWorkshop It is in on Speech and Emotion and Belfast 2000. From the result in such a report, two points that it is about 60% not much poor [(ii) recognition result which the recognition rate of feeling does not change by the difference in (i) language] are clarified.

[0019] When these research results are taken into consideration, although

transfer of the feeling by the meaningless word which does not need transfer of semantics between for example, human being and robot equipment is dramatically ambiguous, it turns out that it is possible. For example, the recognition rate of the feeling is made into about 60%. Moreover, it is shown that it is also possible to compound such utterance by modeling correlation of an emotion and an acoustic feature.

[0020] With the gestalt of operation of this invention, it is made to speak by carrying out the utterance based on such an acoustic feature by making feeling express. Furthermore, with the gestalt of operation of this invention, utterance without (ii) semantics like the (i) speech which is different each time (iii) is realized.

[0021] here, although drawing 1 is a flow chart which shows the basic configuration of the gestalt of operation of the speech-synthesis approach concerning this invention and assumes for example, the robot equipment which has a feeling model, a speech-synthesis means, and a pronunciation means at least as a pronunciation subject, it is limited to this -- not having -- various robot equipments and various computers A.I. Artificial Intelligence other than a robot (artificial intelligence) etc. -- of course, application is also possible In addition, a feeling model is explained in full detail later.

[0022] In this drawing 1 , the feeling condition of a pronunciation subject's feeling

model is distinguished at the first step S1. The condition (feeling condition) of a feeling model changes according to a surrounding environment (external factor) or an internal condition (inner factor), and, specifically, it distinguishes whether it is in ataraxia, the resentment, sadness, joy, composure, and ***** about this feeling condition.

[0023] Here, as a behavioral model, it has the probability state-transition model (for example, model which has a state transition table so that it may mention later) inside, and in the case of robot equipment, it has the transition probability table from which each condition differs with the value of a recognition result, feeling, or instinct, and changes to it to the following condition according to that probability, and the action related with this transition is outputted to it.

[0024] The expressive behavior of the joy and sadness by feeling is described by this probability state-transition model (or probability transition table), and the feeling expression with voice (based on utterance) is included as one of the expressive behavior of this. therefore -- as one element of the action for which it opts when a behavioral model refers to the parameter showing the feeling condition of a feeling model by this example -- a feeling expression -- it is -- a part of action decision section -- distinction of a feeling condition will be performed as a function.

[0025] In addition, this invention is not limited to this example and speech

synthesis which expresses the distinguished feeling condition with voice is performed by step S1 at a next step that distinction of the feeling condition of a feeling model should just be performed at least.

[0026] The utterance sentence which expresses with the following step S2 the content emitted as voice is outputted. This step S2 may be before step S1 or after step S3 mentioned later. Moreover, an utterance sentence new each time may be generated and you may make it choose at random either of two or more utterance sentences which were generated beforehand and prepared. However, it is needed in the gestalt of operation of this invention that it is a meaningless utterance sentence. Although it is difficult to realize the dialogue which had semantics actually when applying to robot equipment, especially this When it can realize with an easy configuration if it is utterance of a meaningless word, and a feeling expression moreover follows It takes into consideration that can make it sensed that it is fully having a dialog also in the meaningless word, and the room of the imagination of the direction of a meaningless word by the side of a hearer spreads, and a sense of closeness and an intimate feeling increase rather than the content of utterance with unsuitable semantics etc. Moreover, by performing generation or selection of an utterance sentence at random, utterance synthesized voice from which and reproduced will come to be different each time, and will always sense fresh.

[0027] Thus, the utterance sentence outputted in this step S2 is a sentence which consists of random words, and, specifically, is realized by making each word into random syllable. the consonant whose syllable here is a phoneme -- it is referred to as valve flow coefficient or conical cup value, combining C and Vowel V. The phoneme is prepared beforehand, and all those phonemes are controlled by the gestalt of operation according to the distinction result of a feeling condition, although it has the persistence time and the pitch as a parameter which were fixed at first. By control of such a parameter according to the distinction result of a feeling condition, it becomes the utterance by which the feeling expression was made. Control of the parameter according to the distinction result of such a feeling expression is explained in full detail later.

[0028] In addition, although the utterance sentence outputted is unrelated to the feeling condition of a feeling model, or its distinction result, it may adjust the utterance sentence outputted to some extent according to a feeling condition etc., or you may make it control generation of an utterance sentence, or the selection processing itself in the gestalt of this operation.

[0029] Next, according to the distinction result of the feeling condition in the above-mentioned step S1, the parameter for speech synthesis is controlled by step S3. The parameters for speech synthesis are things, such as the persistence time of an above-mentioned phoneme, and a pitch or sound volume,

and a feeling expression is made to perform by changing these parameters according to the distinction result of a feeling condition, for example, ataraxia, the resentment, sadness, joy, composure, etc. The parameter combination table corresponding to each feeling (ataraxia, the resentment, sadness, joy, composure, etc.) as an above-mentioned distinction result is created beforehand, and, specifically, switching these tables according to the feeling distinguished actually is mentioned. About the table prepared according to each feeling, the example is shown later.

[0030] In the following step S4, it synthesizes voice from the utterance sentence outputted at the above-mentioned step S2 according to delivery and the parameter controlled by the above-mentioned step S3 to a speech synthesizer (speech synthesizer: speech synthesizer). The obtained voice time series data from which it synthesized voice are emitted as actual voice by being sent to a loudspeaker through a D/A converter, amplifier, etc. For example, when it is robot equipment, such processing is made in the so-called virtual robot, and utterance as which the feeling at that time is expressed comes to be made from a loudspeaker.

[0031] According to the gestalt of fundamental operation of this invention explained above, according to the feeling relevant to a physical condition, by controlling the parameters for speech synthesis (the duration of a phoneme, a

pitch, sound volume, etc.) In order to be able to perform the utterance accompanied by a feeling expression and to choose a phoneme at random, it is necessary to give semantics to neither a word nor the sentence itself. And yet while being able to make the utterance which is different whenever it sounds like utterance, and changes a part of above-mentioned parameter at random or compounds by determining the combination of a phoneme, and the die length of a word and a sentence at random, since there are few parameters to control, mounting is simple.

[0032] (2) Explain a synthetic algorithm with the meaningless word used as feeling, the synthetic algorithm feeling of a meaningless word, and an utterance sentence to a detail. In the gestalt of operation, the place made into the target of a synthetic algorithm is generating the utterance sentence without semantics like a speech which is different each time. And it is making a feeling expression follow on such an utterance sentence.

[0033] A speech synthesizer (speech synthesizer: speechsynthesizer) or a speech synthesis system is used for generation of such an utterance sentence. The inputs of a speech synthesizer (speech synthesizer: speech synthesizer) or speech synthesis cis- TEMUHE are a duration to a list and each phoneme, target pitch, hitting time (it expresses at the percentage to a duration) of a phoneme, etc. The outline of an algorithm of realizing such speech synthesis is

as follows.

[0034] (2-2) Generation of the utterance sentence of the generation meaningless word of an utterance sentence is realized in constituting with a random word. Furthermore, random syllable constitutes each word. the consonant whose syllable is a phoneme here -- it shall come to combine C and Vowel V and shall be expressed as valve flow coefficient or a conical cup value Here, it has the phoneme as a list. And all the phonemes that it has as a list are first registered with the duration and pitch which were fixed.

[0035] For example, a certain phoneme "b" is expressed by the value of "448 10 150 80 158", and is registered into the list. Here, "448" shows that the duration of a phoneme "b" is 448ms. Moreover, it is shown that the following "10" and its following "150" reach 150Hz in 448ms of 10% of persistence time. Moreover, the following "80" and its following "158" show reaching 158Hz to 448ms of 80% of persistence time. Thus, all the phonemes of a list are expressed.

[0036] The syllable expressed by association with the phoneme "b" given by "131 80 179", and the phoneme "@" given by "77 20 200 80 229" and the phoneme "b" given by "405 80 169" is shown in drawing 2 . This example shows as syllable which combined each phoneme which has a discontinuous relation actually so that it might become continuously.

[0037] A feeling expression comes to be made by the utterance sentence

because the phoneme used as what constitutes such syllable can add modification in accordance with each feeling expression. It is changed for a feeling expression of the duration and pitch which are also the information which specifically shows individuality or a property of a phoneme which was mentioned above.

[0038] The utterance sentence constituted from such a phoneme divides roughly, it is the combination of a word and the word is made into the combination of syllable, and further, the syllable is made into the combination of a phoneme and constituted. The processing in <1>- <5> is explained in detail below the whole phase which constitutes such an utterance sentence.

<1> The number of the words in a sentence is decided first. For example, it determines as a random number between 20 - MAXWORDS. MAXWORDS is the maximum number of the word which constitutes a text, and is a parameter for speech synthesis.

<2> Each word is generated. Specifically, it is first determined by the probability (PROBACCENT) whether an accent is in a text at a word.

[0039] The procedure of the following <3-1> - <3-7> determines the syllable of a word, and its phoneme, and a word is determined.

<3-1> The number of the syllable in each word is decided. For example, it determines as a random number between 2 - MAXSYLL. MAXSYLL is the

maximum number of the syllable which constitutes a word, and is a parameter for speech synthesis.

<3-2> Here, if it is a word with an accent, it will choose at random [one] of the syllable, and an accent will be marked.

<3-3> It determines as a thing of valve flow coefficient expression or a conical cup value expression of each syllable. For example, it is determined that the syllable of valve flow coefficient expression is chosen by 0.8% of probability.

<3-4> The consonant and vowel which are assigned to C and V of valve flow coefficient chosen such or conical cup value are read from a phoneme database (or phoneme list) at random.

<3-5> The duration of each phoneme is calculated by $MEANDUR + \text{random}(DURVAR)$. Here, MEANDUR is the persistence time of immobilization and random (DURVAR) is a value determined as a random number. Here, MEANDUR and DURVAR are the parameters for speech synthesis.

<3-6-1> Count of the pitch of a phoneme is calculated by $e = MEANPITCH + \text{random}(PITCHVAR)$. Here, MEANPITCH is the pitch of immobilization and random (PITCHVAR) is a value determined as a random number. Here, MEANPITCH and PITCHVAR are made into the parameter determined according to feeling.

<3-6-2> If a phoneme is a consonant here, the pitch of a consonant is obtained

as e-PITCHVAR. Moreover, if a phoneme is a vowel, a vowel pitch is obtained as e+PITCHVAR.

<3-7-1> If an accent is in syllable, DURVAR will be added at a duration.

<3-7-2> And when there is an accent and it is DEFAULTCONTOUR=rising, the pitch of a consonant is made into MAXPITCH-PITCHVAR and a vowel pitch is made into MAXPITCH+PITCHVAR. Moreover, when it is DEFAULTCONTOUR=falling, the pitch of a consonant is made into MAXPITCH+PITCHVAR and a vowel pitch is made into MAXPITCH-PITCHVAR. Furthermore, when it is DEFAULTCONTOUR= stable, the pitch of a consonant and a vowel is set to MAXPITCH. DEFAULTCONTOUR shows the property (CONTOUR) of syllable. Moreover, MAXPITCH is a parameter for speech synthesis.

[0040] The syllable of a word and its phoneme are determined by the procedure of above <3-1> - <3-7>. And processing which changes contour (profile) of the word which finally comes to the last of an utterance sentence is performed.

<4-1> If there is nothing in the word of the last of a text in an accent, it will be made $e = \text{PITCHVAR}/2$. Here, when it is CONTOURLASTWORD=faling, about each syllable, $-(l+1) * e$ is added and it considers as $e = e + e$. I shows the index of a phoneme. Moreover, when it is parameter CONTOURLASTWORD=rising, about each syllable, $+(l+1) * e$ is added and it considers as $e = e + e$.

<4-2> On the other hand, if it is in the last word in an accent, when it will be CONTOURLASTWORD=falling, DURVAR is added to the duration of each syllable. And the pitch of a consonant is made into MAXPITCH+PITCHVAR and a vowel pitch is made into MAXPITCH-PITCHVAR. When it is parameter CONTOURLASTWORD=rising, DURVAR is added to the duration of each syllable. And the pitch of a consonant is made into MAXPITCH-PITCHVAR and a vowel pitch is made into MAXPITCH+PITCHVAR.

<5> Finally the volume of the whole sentence is set as VOLUME. Here, VOLUME is a parameter for speech synthesis.

[0041] An utterance sentence comes to be generated by the processing in every above phases. And since some parameters used for the decision of an utterance sentence are given with the random number, the meaningless word without semantics which is different each time can be generated. And a feeling expression comes to be made by such utterance sentence by giving various parameters which were mentioned above according to feeling.

[0042] In addition, description of a program (source code) for hardware to realize processing which was mentioned above is shown in drawing 3 and drawing 4 . A part for the pre-stage of a program is shown in drawing 3 , and a part for a post-stage is shown in drawing 4 .

[0043] (2-2) A feeling expression can be given to an utterance sentence by

controlling the parameter which is given according to each feeling and which was used in the algorithm of text generation which carried out parameter **** according to feeling. Here, as feeling that an utterance sentence is expressed such for example, ataraxia, the resentment, sadness, joy, or composure (calm, anger, sadness, happiness, comfort) is mentioned. In addition, it cannot be overemphasized that it is not limited to the feeling listed here.

[0044] For example, such an emotion can express rousing (Arousal) and a potency (valence) on the feature space used as an element, respectively. for example, it comes out as are shown in drawing 5 , and the resentment, sadness, and the field in which it is glad or settles (anger, sadness, happiness, comfort) are constituted on the feature space which uses rousing (Arousal) and a potency (valence) as an element and the field of ataraxia (calm) is constituted by the core. for example, "the resentment (anger)" is expressed as rousing as negative -- having -- "-- feeling sad (sadness) -- " -- if it is not "rousing, it is expressed as negative."

[0045] The combination (at least duration (DUR), pitch (PITCH), sound volume (VOLUME), etc. of phoneme) table of the parameter beforehand determined corresponding to each feeling, such as resentment, sadness, joy, and composure, is shown in the following tables. Such a table is beforehand generated based on the special feature of each feeling.

[0046]

[A table 1]

[0047]

[A table 2]

[0048]

[A table 3]

[0049]

[A table 4]

[0050]

[A table 5]

[0051] Thus, control according to the feeling of the parameter used for speech synthesis is realized by switching the table which consists of a parameter corresponding to each feeling prepared beforehand according to the feeling distinguished actually.

[0052] And the utterance sentence by which the feeling expression was made comes to be generated by making the speech synthesis which used the parameter of the table chosen according to feeling such. And although human being does not understand the content of utterance of robot equipment itself when robot equipment utters the utterance sentence of a meaningless word by which the feeling expression generated in this way was made, the feeling of robot equipment can be known. And since such utterance becomes what is further different each time, human being can always sense a dialogue with robot equipment fresh. Next, the robot equipment which is the gestalt of operation of this invention is explained, and the mounting gestalt of the algorithm of

above-mentioned utterance to such robot equipment is explained concretely after that.

[0053] In addition, although it has realized with the gestalt of operation by switching the table which consists of a parameter which has prepared control of the parameter according to feeling beforehand. corresponding to feeling by actual feeling, it cannot be overemphasized that control of the parameter according to feeling is not limited to the gestalt of this operation.

[0054] (3) Explain to a detail the example which applied this invention to the autonomous mold pet robot of 4 guide pegs as a gestalt of more concrete operation of this invention below the configuration of the example (3-1) robot equipment of the robot equipment by the gestalt of this operation, referring to a drawing. Feeling and an instinct model are introduced into the software of this pet mold robot equipment, and it can be made to do [obtaining action more near a living thing, or]. Although the robot which operates actually is used with the gestalt of this operation, if the utterance by the meaningless word is computer system with a loudspeaker, it can realize easily and it is an effective function in the field of the interaction (or dialogue) of human being and a machine. Therefore, there is no applicability of this invention what is restricted to a robot system.

[0055] The head unit 4 and the tail section unit 5 are connected with the front

end section and the back end section of the idiosoma unit 2, respectively, and the robot equipment as an example is constituted while considering as the so-called pet robot of the configuration which imitated the "dog" and connecting the leg units 3A and 3B, 3C, and 3D with front and rear, right and left of the idiosoma unit 2, respectively, as shown in drawing 6 .

[0056] As shown in the idiosoma unit 2 at drawing 7 , the control section 16 formed by connecting CPU (Central Processing Unit)10, DRAM (Dynamic Random Access Memory)11, a flash ROM (Read Only Memory) 12, PC (Personal Computer) card interface circuitry 13, and a digital disposal circuit 14 mutually through an internal bus 15 and the dc-battery 17 as a source of power of this robot equipment 1 are contained. Moreover, the angular-velocity sensor 18, an acceleration sensor 19, etc. for detecting the sense of robot equipment 1 and the acceleration of a motion are contained by the idiosoma unit 2.

[0057] Moreover, the CCD (Charge Coupled Device) camera 20 for picturizing an external situation to the head unit 4, The touch sensor 21 for "it strokes" and the physical influence of "striking" from a user to detect a carrier beam pressure, The distance robot 22 for measuring the distance to the body located ahead, LED (Light Emitting Diode) (not shown) equivalent to the microphone 23 for collecting alien frequencies, the loudspeaker 24 for outputting voice, such as a cry, and the "eye" of robot equipment 1 etc. is arranged in the predetermined

location, respectively.

[0058] Furthermore, Actuators 251-25n and Potentiometers 261-26n for free frequency are arranged in the joining segment of the joint part of each leg unit 3A-3D, each joining segment of each leg unit 3A-3D and the idiosoma unit 2, the head unit 4, and the idiosoma unit 2, and the list by the joining segment of tail 5A of the tail section unit 5, respectively. For example, Actuators 251-25n have the servo motor as a configuration. Leg unit 3A - 3D are controlled by actuation of a servo motor, and it changes in a target position or actuation.

[0059] And LED and 251-25n of each actuator are connected with the digital disposal circuit 14 of the control section 16 through the hubs 271-27n corresponding to various sensor lists, such as these angular-velocity sensor 18, an acceleration sensor 19, a touch sensor 21, a distance robot 22, a microphone 23, a loudspeaker 24, and each potentiometers 261-26n, respectively, and direct continuation of CCD camera 20 and the dc-battery 17 is carried out to the digital disposal circuit 14, respectively.

[0060] A digital disposal circuit 14 incorporates sensor data, and the image data and voice data which are supplied from each above-mentioned sensor one by one, and carries out sequential storing of these through an internal bus 15 in the predetermined location in DRAM11, respectively. Moreover, a digital disposal circuit 14 incorporates the dc-battery residue data showing the dc-battery

residue supplied from a dc-battery 17 with this one by one, and stores this in the predetermined location in DRAM11.

[0061] Thus, each sensor data stored in DRAM11, image data, voice data, and dc-battery residue data are used in case CPU10 performs motion control of this robot equipment 1 after this.

[0062] In practice, at the time of the first stage when the power source of robot equipment 1 was switched on, CPU10 reads directly the control program stored in the memory card 28 or flash ROM 12 with which the PC Card slot which the idiosoma unit 2 does not illustrate was loaded through the PC card interface circuitry 13, and stores this in DRAM11.

[0063] Moreover, CPU10 judges [after this] the situation of self and a perimeter, the existence of the directions from a user, and influence, etc. from a digital disposal circuit 14 based on each sensor data by which sequential storing is carried out, image data, voice data, and dc-battery residue data to DRAM11 as mentioned above.

[0064] Furthermore, CPU10 can make the head unit 4 able to shake vertically and horizontally, can move tail 5A of the tail section unit 5, or makes it act by making the required actuators 251-25n drive based on the decision result concerned to make it walk by making each leg unit 3A-3D drive etc. while it opts for the action which stores in this decision result and DRAM11, and continues

based on *****.

[0065] Moreover, in this case, CPU10 generates voice data if needed, by giving this to a loudspeaker 24 as a sound signal through a digital disposal circuit 14, makes the voice based on the sound signal concerned output outside, or turns on, switches off or blinks above-mentioned LED.

[0066] Thus, in this robot equipment 1, it is made as [act / according to the situation of self and a perimeter the directions from a user, and influence / it / autonomously].

[0067] (3-2) the software configuration of a control program -- here comes to show the software configuration of the above-mentioned control program in robot equipment 1 to drawing 8 . In this drawing 8 , the device driver layer 30 is located in the lowest layer of this control program, and consists of device driver sets 31 which consist of two or more device drivers. In this case, each device driver is the object allowed ** which carries out direct access to the hardware used by usual computers, such as CCD camera 20 (drawing 7) and a timer, and processes in response to interruption from corresponding hardware.

[0068] Moreover, the ROBOTIKKU server object 32 With the virtual robot 33 which becomes by the software group which offers the interface for being located in the lowest layer of the device driver layer 30, for example, accessing hardware, such as various above-mentioned sensors and Actuators 251-25n

With the BAWA manager 34 who becomes by the software group which manages the change of a power source etc. It consists of a device driver manager 35 who becomes by the software group which manages other various device drivers, and a dither INDO robot 36 which becomes by the software group which manages the device of robot equipment 1.

[0069] The manager object 37 consists of an object manager 38 and a service manager 39. The object manager 38 is a software group which manages starting and termination of each software group contained in the ROBOTIKKU server object 32, the middleware layer 40, and the application layer 41, and a service manager 39 is a software group which manages connection of each object based on the initial entry between each object described by the connection file stored in the memory card 28 (drawing 7).

[0070] The middleware layer 40 is located in the upper layer of the ROBOTIKKU server object 32, and consists of software groups which offer the fundamental function of these robot equipments 1, such as an image processing and speech processing. Moreover, the application layer 41 is located in the upper layer of the middleware layer 40, and consists of software groups for opting for action of robot equipment 1 based on the processing result processed by each software group which constitutes the middleware layer 40 concerned.

[0071] In addition, the concrete software configuration of the middleware layer

40 and the application layer 41 is shown in drawing 9 , respectively.

[0072] As shown in drawing 9 , the middleware layer 40 The object for noise detection, for temperature detection, The recognition system 60 which has the input semantics converter module 59 etc. in the object for brightness detection, the object for scale recognition, the object for distance detection, the object for position detection, the object for touch sensors, the object for motion detection, and each signal conditioning module 50 for color recognition - 58 lists, It consists of output systems 69 which have each signal conditioning modules 61-67 for the object for position management, the object for tracking, the object for motion playback, the object for a walk, the object for a fall return, the object for LED burning, and sound playback etc. in output semantics converter module 68 list.

[0073] Each signal conditioning modules 50-58 of the recognition system 60 incorporate the data with which it corresponds of each sensor data by which reading appearance is carried out from DRAM11 (drawing 7) with the virtual robot 33 of the ROBOTIKKU server object 32, or image data and voice data, perform predetermined processing based on the data concerned, and give a processing result to the input semantics converter module 59. Here, the virtual robot 33 is constituted by the predetermined protocol as a part which carries out transfer or conversion of a signal.

[0074] The input semantics converter module 59 It is based on the processing

result given from each [these] signal conditioning modules 50-58. "The fall was detected", ["it is "noisy", hot / "hot" /, and bright", "the ball having been detected", and] The self of "it was stroked", "it having been struck", "the scale of C-E-G having been heard", "the body which moves having been detected", "having detected the obstruction", etc. and a surrounding situation, the command from a user, and influence are recognized, and a recognition result is outputted to the application layer 41 (drawing 7).

[0075] Application layer 41., as shown in drawing 10 , it consists of five modules, the behavioral model library 70, the action change module 71, the study module 72, the feeling model 73, and the instinct model 74. Here, the feeling model 73 is a model to which the condition of the feeling that a condition changes with the stimuli from the outside etc. is changed, and superposition of the feeling expression to an utterance sentence which was mentioned above according to the feeling determined with such a feeling model is made. Moreover, the monitor of such a condition of the feeling model 73 or instinct model 74 grade, i.e., distinction of the condition etc., is made by the control means of CPU10 grade.

[0076] As shown in drawing 11 , "when a dc-battery residue decreases", "when avoiding an obstruction, and expressing feeling", the behavioral model library 70 is made to correspond to the condition item of the shoes "at the time of detecting a ball" etc. chosen beforehand, respectively, and the behavioral models 701-70n

which became independent, respectively "are formed ["a fall return is carried out" and] in it."

[0077] And these behavioral models 701-70n The time of a recognition result being given from the input semantics converter module 59, respectively, The parameter value of the corresponding emotion currently held like the after-mentioned at the feeling model 73 if needed when fixed time amount has passed, after the last recognition result is given, It opts for the action which continues while referring to the parameter value of the corresponding desire currently held at the instinct model 74, respectively, and a decision result is outputted to the action change module 71.

[0078] In the case of the gestalt of this operation, in addition, each behavioral models 701-70n As the technique of opting for the next action As opposed to the arc ARC1 - ARCn which connect between to each node NODE0 - NODEn for to other nodes NODE0 of which - NODEn it changes from one node (condition) NODE0 as shown in drawing 12 - NODEn The algorithm called the finite stochastic automaton determined probable based on the transition probability P1-Pn set up, respectively is used.

[0079] Concretely, each behavioral models 701-70n are made to correspond to the node NODE0 which forms the self behavioral models 701-70n, respectively - NODEn, respectively, and have the state transition table 80 as shown in drawing

13 for every these node NODE0 - NODEn.

[0080] In this state transition table 80, the input event (recognition result) made into transition conditions in that node NODE0 - NODEn is listed by the line of an "input event name" at a priority, and the further conditions about that transition condition are described by the "data name" and the corresponding train in the line of the "data range."

[0081] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 13 When the recognition result of "detecting a ball (BALL)" is given The range of "magnitude (SIZE)" of the ball given with the recognition result concerned is "0 to 1000", When the recognition result of "detecting an obstruction (OBSTACLE)" is given, they have been conditions for that the range of "the distance (DISTANCE)" to the obstruction done with the recognition result concerned is "0 to 100" to change to other nodes.

[0082] Moreover, in this node NODE100, when there is no input of a recognition result, it also sets. The inside of each emotion held at the feeling model 73 and the instinct model 74 which 701-70n of behavioral models refers to periodically, respectively, and the parameter value of each desire, it was held at the feeling model 73 -- "-- glad (JOY) -- " -- "-- surprised (SURPRISE) -- " -- or -- "-- feeling sad (SUDNESS) -- " -- when the range of one of parameter value is "50 to 100", it can change to other nodes.

[0083] moreover -- a state transition table 80 -- "-- others, while the node name which can change from the node NODE0 - NODEn in the train of the "transition place node" in the column of transition probability" of NODOHE is listed It is described by the part where it corresponds in the column of transition probability" of NODOHE, respectively. the transition probability to each of other node NODE0 which can change when all the conditions described by the line of an "input event name", a "data value", and the "range of data" are met - NODEn -- "-- others -- the action which should be outputted in case it changes to the node NODE0 - NODEn -- "-- others -- it is described by the line of "output action" in the column of transition probability" of NODOHE. in addition -- "-- others -- the sum of the probability of each line in the column of transition probability" of NODOHE is 100 [%].

[0084] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 13 for example, when the recognition result that it carries out "detecting a ball (BALL)", and the range of "SIZE (magnitude)" of the ball is "0 to 1000" is given It can change to "a node NODE120 (node 120)" by the probability of "30 [%]", and action of "ACTION1" will be then outputted.

[0085] When they are constituted as a lot of nodes NODE0 described as such [respectively] a state transition table 80 - NODEn(s) are connected, and a recognition result is given from the input semantics converter module 59, each

behavioral models 701-70n opt for the next action probable using the state transition table of the node NODE0 - NODEn, and are made as [output / to the action change module 71 / a decision result].

[0086] The action change module 71 shown in drawing 10 chooses from each behavioral models 701-70n of the behavioral model library 70 the action outputted from the high behavioral models 701-70n of the priority beforehand defined among the actions outputted, respectively, and sends out the command (this is hereafter called action command.) of the purport which should perform the action concerned to the output semantics converter module 68 of the middleware layer 40. In addition, in the gestalt of this operation, priority is set up highly about 701-70n of behavioral models written by the bottom in drawing 11 .

[0087] Moreover, the action change module 71 notifies that the action was completed based on the completion information of action given from the output semantics converter module 68 after the completion of action to the study module 72, the feeling model 73, and the instinct model 74.

[0088] On the other hand, the recognition result of carrier beam instruction is inputted as influence from a user, such as the study module 72 "was struck" among the recognition results given from the input semantics converter module 59, and "it having been stroked."

[0089] And based on the advice from this recognition result and the action

change module 71, the study module 72 reduces the manifestation probability of that action, when "struck" (scolded), and when "stroked" (praised), it changes corresponding behavioral models [in the behavioral model library 70 / 701-70n] corresponding transition probability so that the manifestation probability of that action may be raised.

[0090] on the other hand, the feeling model 73 -- "-- glad (joy) -- " -- "-- feeling sad (sadness) -- " -- "the resentment (anger)" -- "-- surprised (surprise) -- " -- "dislike (disgust)" -- and -- "-- afraid (fear) -- " -- the parameter with which the strength of the emotion is expressed for every emotion is held about a total of six emotions. And the feeling model 73 updates the parameter value of each [these] emotion periodically based on the advice from the specific recognition result to which it is given from the input semantics converter module 59, respectively, such as "it having been struck" and "it having been stroked", and elapsed time and the action change module 71 etc.

[0091] The recognition result to which the feeling model 73 is specifically given from the input semantics converter module 59, The amount of fluctuation of the action and its emotion when being computed by predetermined operation expression based on the elapsed time after updating last time etc. of the robot equipment 1 at that time $E[t]$, The multiplier which expresses the sensibility of $E[t]$ and its emotion for the parameter value of the current emotion is set to k_e .

(1) By the formula, as parameter value [of the emotion in the following period] $E[t+1]$ is computed and this is replaced with parameter value [of the current emotion] $E[t]$, the parameter value of the emotion is updated. Moreover, the feeling model 73 updates the parameter value of all emotions like this.

[0092]

[Equation 1]

[0093] In addition, it is decided beforehand what the advice from each recognition result or the output semantics converter module 68 has effect of on amount of fluctuation α [of the parameter value of each emotion] $E[t]$. For example, the recognition result of "having been struck" has big effect on amount of fluctuation α [of the parameter value of the emotion of the "resentment"] $E[t]$, and the recognition result of "having been stroked" has big effect on amount of fluctuation α [of the parameter value of the emotion of "joy"] $E[t]$.

[0094] Here, the advice from the output semantics converter module 68 is the so-called feedback information (the completion information of action) of action, and is the information on the appearance result of action, and the feeling model 73 changes feeling also using such information. this -- for example, the feeling level of the resentment falls by action of "barking" -- like -- they are things. In

addition, the advice from the output semantics converter module 68 is inputted also into the study module 72 mentioned above, and the study module 72 changes behavioral models [701-70n] corresponding transition probability based on the advice.

[0095] In addition, feedback of an action result may be made with the output (action to which feeling was added) of the action change modulator 71.

[0096] on the other hand, "motion avarice (exercise)", "love avarice (affection)", "appetite (appetite)", and "the curiosity (curiosity) of the instinct model 74" are mutually-independent -- the parameter with which the strength of the desire is expressed for these the desires of every is held about four desires the bottom. And the instinct model 74 updates the parameter value of these desires periodically based on the recognition result to which it is given from the input semantics converter module 59, respectively, the advice from elapsed time and the action change module 71, etc.

[0097] The instinct model 74 specifically about "motion avarice", "love avarice", and "curiosity" The amount of fluctuation of that the desire of the when being computed by predetermined operation expression based on the advice from a recognition result, elapsed time, and the output semantics converter module 68 etc. $\Delta I[k]$, The parameter value of the current desire as a multiplier k_i showing the sensibility of $I[k]$ and its desire As parameter value [of that desire in

the following period $I[k+1]$ is computed using (2) types with a predetermined period and this result of an operation is replaced with parameter value I of that current desire $I[k]$, the parameter value of that desire is updated. Moreover, the instinct model 74 updates the parameter value of each desire except "appetite" like this.

[0098]

[Equation 2]

[0099] In addition, it is decided beforehand what the advice from a recognition result and the output semantics converter module 68 etc. has effect of on amount of fluctuation Δ of the parameter value of each desire $I[k]$, for example, it has effect to amount of fluctuation Δ of the parameter value of the "fatigue" $I[k]$ with the big advice from the output semantics converter module 68.

[0100] In addition, in the gestalt of this operation, it is regulated so that each emotion and the parameter value of each desire (instinct) may be changed in the range from 0 to 100, respectively, and the value of multipliers k_e and k_i is also set up according to the individual for each I (an emotion and I the desire of every.

[0101] On the other hand, abstract action commands, such as it being I which is given from the action change module 71 of the application layer 41 as mentioned

above] "advance", "it being glad", the output semantics converter module 68 of the middleware layer 40 "cries", as shown in drawing 9 , or "tracking (a ball is pursued)", are given to the signal conditioning modules 61-67 with which the output system 69 corresponds.

[0102] And these signal conditioning modules 61-67 The servo command value which should be given to the actuators 251-25n (drawing 7) in order to perform the action based on the action command concerned, if an action command is given, Or the actuation data given to LED of a "eye" are generated. the voice data of the sound outputted from a loudspeaker 24 (drawing 7) -- and -- Sequential sending out of these data is carried out at the actuators 251-25n which correspond through the virtual robot 33 and digital disposal circuit 14 (drawing 7) of the ROBOTIKKU server object 32 one by one, a loudspeaker 24, or LED.

[0103] Thus, in robot equipment 1, it is made based on the control program as [perform / the situation of self (interior) and a perimeter (exterior), the directions from a user, and autonomous action according to influence].

[0104] (3-3) Robot equipment can be constituted as the algorithm of the utterance to robot equipment carried out mounting ****. Argo RISUMU of above-mentioned utterance is mounted as a sound playback module 67 in drawing 9 of such robot equipment 1.

[0105] By the sound playback module 67, the sound output command (for example, "speak with joy" etc.) determined in the part (for example, behavioral model) of a high order is received, actual voice time series data are generated, and data are transmitted to the virtual robot's 33 loudspeaker device in order. The utterance sentence which consists of a meaningless word in which the feeling expression was made from the loudspeaker 24 shown in drawing 7 in robot equipment by this is emitted.

[0106] The behavioral model (henceforth an utterance behavioral model) which generates the utterance command doubled with feeling is explained. The utterance behavioral model is prepared as a behavioral model of 1 in the behavioral model library 70 shown in drawing 10 .

[0107] In the utterance behavioral model, the content of utterance is determined using the state transition table 80 as always shown in drawing 10 with reference to the newest parameter value based on such each parameter value from the feeling model 73 or the instinct model 74. Namely, it is made to perform utterance action adapted to the feeling at the time of transition, using the value of feeling as transition conditions from a certain condition.

[0108] The state transition table which an utterance behavioral model uses can be expressed as shown in drawing 14 . In addition, the state transition tables used for the utterance behavioral model shown in drawing 14 do not differ

substantially, although notation formation of the state transition table 80 shown in above-mentioned drawing 13 differs. The state transition table shown like drawing 14 is explained.

[0109] this example -- a node -- "node XXX" from -- it is glad as transition conditions to other nodes (HAPPY), and feels sad (SAD), and the resentment (ANGER) and a time-out (TIMEOUT) are given. and -- glad (HAPPY) -- feeling sad (SAD) -- the concrete numeric value as the resentment (ANGER) and transition conditions to a time-out (TIMEOUT) -- respectively -- It is given as HAPPY>70, SAD>70, ANGER>70, and TIMEOUT=timeout.1. Here, timeout.1 is a numeric value, for example, it is a value which shows predetermined time.

[0110] moreover, the action which nodeYYY, nodeZZZ, nodeWWW, and nodeVVV are prepared and is performed to such each node as a node to which a node can change from "node XXX" -- respectively -- "BANZAI (BANZAI)" -- "it falling (OTIKOMU)" -- "-- shiveringly (BURUBURU) -- " -- and it is assigned as "a yawn (AKUBI)."

[0111] Here, the expressive behavior of "Banzai! (BANZAI)" is defined as what carries out the utterance (talk_happy) by which the feeling expression of "the joy" is carried out, and carries out actuation (motion_banzai) of the banzai by the nose gear etc., and carries out actuation (motion_swingtail) which wags a tail further. In order to carry out the utterance which carried out the feeling

expression of "joy" here, the parameter for a feeling expression of the joy currently prepared beforehand which was mentioned above is used. That is, joy is uttered based on the algorithm of utterance currently explained previously.

[0112] Moreover, the expressive behavior of "falling (OTIKOMU)" is defined as what carries out the utterance (talk_sad) by which the feeling expression of "the sadness" is carried out, and carries out the so-called timid actuation (motion_ljijji). In order to carry out the utterance which carried out the feeling expression of "sadness" here, the parameter for a feeling expression of the sadness currently prepared beforehand which was mentioned above is used. That is, sadness is uttered based on the algorithm of utterance currently explained previously.

[0113] moreover -- "-- shiveringly (BURUBURU) -- " -- expressive behavior is defined as what carries out actuation (motion_buruburu) which carries out the utterance (talk_anger) by which the feeling expression of the "resentment" is carried out, and is trembling for the resentment. In order to carry out the utterance which carried out the feeling expression of the "resentment" here, the parameter for the feeling expression of resentment which was mentioned above and which is prepared beforehand is used. That is, the resentment is uttered based on the algorithm of utterance currently explained previously.

[0114] Moreover, the expressive behavior of "a yawn (AKUBI)" is defined as

actuation (motion_akubi) which yawns since it is tedious without nothing.

[0115] Thus, each action performed in each node which can change is defined, and the transition to such each node is determined by the probability table. That is, the probability table on which the action probability at the time of agreeing on transition conditions was described has determined the transition to each node.

[0116] In the example shown in drawing 14 , when glad (HAPPY) (i.e., when the value of HAPPY exceeds 70 made into a predetermined threshold), it is chosen by the expressive behavior of "Banzai! (BANZAI)" by 100% of probability. Moreover, when feeling sad (i.e., when the value of SAD exceeds 70 made into a predetermined threshold) (SAD), it is chosen by the expressive behavior of "falling (OTIKOMU)" by 100% of probability. moreover -- the case where the value of the **** rate ANGER exceeds 70 made into a predetermined threshold in the case of the resentment (ANGER) -- 100% of probability -- "-- shiveringly (BURUBURU) -- " -- expressive behavior is chosen. And in the case of a time-out (TIMEOUT) (i.e., when the value of TIMEOUT is set to timeout.1 made into a predetermined threshold), the expressive behavior of "a yawn (AKUBI)" is chosen by 100% of probability. in addition -- although the case where 100% of probability shows altogether the case where action is chosen, namely, action surely speaks is made into the example in this example -- this -- definition -- ** -- there are nothings. That is, when glad (HAPPY), you may make it, choose action

of "Banzai! (BANZAI)" at 70% for example.

[0117] By defining the state transition table of an utterance behavioral model as mentioned above, it comes be made [controlling the utterance corresponding to feeling freely according to the input of other sensors, or a robot's condition, or] to robot equipment.

[0118] In addition, as a parameter controlled by feeling, the persistence time, a pitch, and sound volume were mentioned as the example, and the gestalt of above-mentioned operation explained them. However, the text configuration factor influenced by feeling can also be used as a parameter, without being limited to this.

[0119] Moreover, by explanation of the gestalt of above-mentioned operation, the feeling model of robot equipment was glad and the case where it was constituted by feeling, such as resentment, was explained. However, it is not limited to a feeling model being constituted by the feeling mentioned as such an example, and other factors which affect feeling can also constitute. And the parameter which constitutes a text is controlled by such other factors in this case.

[0120]

[Effect of the Invention] The feeling distinction process that the speech synthesis approach concerning this invention distinguishes the feeling condition of a pronunciation subject's feeling model, The utterance sentence output process

which outputs the utterance sentence showing the content emitted as voice, and the parameter-control process which controls the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction process, By having the speech synthesis process which synthesizes voice based on the parameter which inputted into the speech synthesis section the utterance sentence outputted by the utterance sentence output process, and was controlled Based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model, a pronunciation subject's utterance sentence is generable.

[0121] Moreover, a feeling distinction means by which the voice synthesizer concerning this invention distinguishes the feeling condition of a pronunciation subject's feeling model, An utterance sentence output means to output the utterance sentence showing the content emitted as voice, and a parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction means, By having the speech synthesis means from which the utterance sentence outputted by the utterance sentence output means is supplied, and synthesizes voice based on the controlled parameter According to the feeling condition distinguished by feeling distinction means to distinguish the feeling condition of a pronunciation subject's feeling model, the parameter for speech synthesis is

controlled by the parameter-control means. The utterance sentence outputted by the utterance sentence output means is supplied, and can synthesize voice with a speech synthesis means based on the controlled parameter. Thereby, a voice synthesizer can generate a pronunciation subject's utterance sentence based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model.

[0122] Moreover, the feeling model with which the robot equipment concerning this invention originates in actuation, A feeling distinction means to distinguish the feeling condition of a feeling model, and an utterance sentence output means to output the utterance sentence showing the content emitted as voice, A parameter-control means to control the parameter for speech synthesis according to the feeling condition distinguished by the feeling distinction means, By having the speech synthesis means from which the utterance sentence outputted by the utterance sentence output means is supplied, and synthesizes voice based on the controlled parameter According to the feeling condition distinguished by feeling distinction means to distinguish the feeling condition of the feeling model resulting from actuation, the parameter for speech synthesis is controlled by the parameter-control means. The utterance sentence outputted by the utterance sentence output means is supplied, and can synthesize voice with a speech synthesis means based on the controlled parameter. Thereby, robot

equipment can generate a pronunciation subject's utterance sentence based on the parameter for the speech synthesis controlled according to the feeling condition of an utterance subject's feeling model.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the flow chart which shows the basic configuration of the gestalt of operation of the speech synthesis approach concerning this invention.

[Drawing 2] It is drawing showing the relation between the persistence time of each phoneme, and a pitch.

[Drawing 3] It is a program for creation of the utterance sentence by speech synthesis, and is drawing showing a part for the first portion.

[Drawing 4] It is a program for creation of the utterance sentence by speech synthesis, and is drawing showing a part the second half.

[Drawing 5] It is drawing showing the relation of the class of each feeling in a feature space or a plane of action.

[Drawing 6] It is the perspective view showing the appearance configuration of the robot equipment which is the gestalt of operation of this invention.

[Drawing 7] It is the block diagram showing the circuitry of above-mentioned robot equipment.

[Drawing 8] It is the block diagram showing the software configuration of above-mentioned robot equipment.

[Drawing 9] It is the block diagram showing the configuration of the middleware layer in the software configuration of above-mentioned robot equipment.

[Drawing 10] It is the block diagram showing the configuration of the application layer in the software configuration of above-mentioned robot equipment.

[Drawing 11] It is the block diagram showing the configuration of the behavioral model library of an above-mentioned application layer.

[Drawing 12] It is drawing used in order to explain the finite stochastic automaton used as the information for the action decision of robot equipment.

[Drawing 13] It is drawing showing the state transition table prepared for each node of a finite stochastic automaton.

[Drawing 14] It is drawing showing the state transition table of an utterance behavioral model.

[Description of Notations]

1 Robot Equipment, 10 CPU, 14 Digital Disposal Circuit, 24 Loudspeaker, 70 Behavioral Model, 73 Feeling Model